TIGHT BINDING BOOK

INTRODUCTION

TO

THE THEORY

OF

ANALYTIC FUNCTIONS

BY

PROFESSOR JAMES HARKNESS

AND

PROFESSOR FRANK MORLEY

REPRINT 1924.

G. E. STECHERT & Co.

NEW-YORK

PREFACE.

THE present book is not to be regarded as an abridged and more elementary version of our treatise on the theory of functions, but as an independent work. It has been composed with different ends in mind, deals in many places with distinct orders of ideas, and presents from an independent point of view such portions of the subject-matter as are common to both volumes. In the treatise our desire was to cover as fully as possible within the limits at our disposal a very extensive field of analysis, and the execution of this plan precluded the possibility of allotting much space to preliminary notions. At the same time we recognized that readers approaching the subject for the first time could not fail to be hampered by the non-existence in English of any text-book giving a consecutive and elementary account of the fundamental concepts and processes employed in the theory of functions; subsequent experience and inquiry has only strengthened our belief that if English and American students are not to be placed at a disadvantage as compared with their foreign brethren they should have ready access to text-books which discuss topics of the kind indicated above. It was to an attempt to meet these requirements to the best of our ability that the present volume owes its genesis, and this is its bond of connexion with the treatise.

The theory of functions, by virtue of its immense range and vitality and its innumerable points of contact with other branches

of mathematics, has taken a central position in modern analysis, and has made its influence felt in all parts of the mathematical domain. It is not surprising accordingly to find that such of the current text-books as have been composed in the modern spirit show numerous traces of the inrush of new ideas due to a wider acquaintance with the theory of functions, and that they are, both as regards structure and aim, poles apart from those of the preceding generation. There is, however, much still to be done in the direction of recasting elementary mathematics in the light of recent knowledge; in particular works in English that treat of the scientific parts of arithmetic show little if any trace of recent discoveries with respect to the number-system. As we felt that it would be unsafe to assume any acquaintance with the various modern views on the nature of ordinal and cardinal numbers, and as it was indispensable for the proper comprehension of the succeeding chapters that the meaning of the term ordinal number should be clearly appreciated, we have devoted the first chapter to the discussion of what is meant by an ordinal number. This chapter is not and lays no claim to being a scientifically complete account of the matter; it will serve its object if it conveys to the reader a distinct image of a number divorced from measurement.

In places we have gone afresh over old ground; this has been done either for the sake of organic unity, or in order to emphasize by means of simple examples ideas which appear later in more difficult and complicated forms. It has in fact been our desire to keep the difficulties of the subject apart from those which are merely difficulties of technique. In carrying out this plan we have consistently chosen the simplest available examples.

As regards the theory of functions proper we had to make a choice between the methods of Cauchy and those of Weierstrass. While fully alive to the wonderful beauty and power of Cauchy's theory, we decided eventually in favour of Weierstrass's system. Weierstrass has himself stated with his usual lucidity and force the reasons which have led him to

prefer his own scheme to that of Cauchy and Riemann for the purposes of a systematic construction of a theory of functions. In a letter to Prof. Schwarz (Weierstrass, *Ges. Werke,* vol. ii. p. 235) he says :—"The more I reflect upon the principles of the theory of functions,—and I do so incessantly,—the stronger becomes my conviction that this theory must be built up on the foundation of algebraic truths and that therefore it is not the right way if we proceed conversely and call into play the *transcendental* (to express myself briefly) in order to establish simple and fundamental algebraic theorems,—however attractive, for example, the considerations may be by which Riemann discovered so many of the most important properties of algebraic functions. It is self-evident that all routes ought to be free to the investigator, while engaged on his researches; I am thinking only of the systematic establishment of the theory." It has seemed to us that for the purposes of an *introductory* work it was important to secure the advantages of homogeneity, intrinsic logical consistency, and the gradual passage from the simple to the complex in place of the reverse, which form so marked a feature of Weierstrass's system. With this in mind we have, in the main, followed Weierstrassian lines. It would, however, have been mere pedantry to exclude all geometric considerations from a book intended for the use of beginners. This is the justification,—if any justification should be thought necessary,— of the geometric chapters.

The bilinear transformation has been discussed in much detail. This transformation is interesting in itself and can be used effectively to bring out many points of importance for the general theory; furthermore it is playing an increasingly prominent part in recent mathematical work and a complete mastery of its properties is now an indispensable prerequisite for the study of the more advanced portions of our subject.

Without entering into minute particulars with respect to the remaining chapters we may indicate in a few words the general principles which have guided us in our choice of materials. We have kept steadily in view the desirability of

making the book elementary, wherever that was possible without any sacrifice of thoroughness; we have sought to remove erroneous notions which we have found prevalent among beginners; we have tried to avoid a one-sided development and to bring home to the mind of the student the vital significance of theorems which may appear wholly abstract by using such concrete illustrations as elliptic functions, definite integrals, the potential, etc.; and finally we have excluded, intentionally, all theorems and results which, however beautiful intrinsically, seem unlikely to be of assistance to the student who proposes to carry his studies further. If we should prove to have met with some measure of success in carrying out this arduous programme, we shall be amply rewarded.

We have not given many references, partly because this has been done very fully in our larger book, partly also because we have used our material in a way which differs widely from that employed in other books.

Mr Arthur Berry, M.A., Fellow and Assistant Tutor of King's College, Cambridge, has been untiring in his assistance in the revision of the proof-sheets. His great knowledge of the subject and keen critical insight have enabled us to make many improvements both in substance and form, and in numberless ways his advice has been of the greatest value to us. To other friends who have taken an interest in the progress of this work we desire to express our sense of gratitude.

Finally we must thank the Macmillan Company and the Officers of the Cambridge University Press for the help that they have given to us in the publication of this volume, and for their admirable efficiency in the preparation and printing.

J. HARKNESS.
F. MORLEY.

June 11, 1898.

CONTENTS.

CHAPTER I.

THE ORDINAL NUMBER-SYSTEM.

CHAPTER II.

THE GEOMETRIC REPRESENTATION OF COMPLEX NUMBERS.

CHAPTER III.

THE BILINEAR TRANSFORMATION.

CHAPTER IV.

GEOMETRIC THEORY OF THE LOGARITHM AND THE EXPONENTIAL.

CHAPTER V.

THE BILINEAR TRANSFORMATION OF A PLANE INTO ITSELF.

CHAPTER VI.

LIMITS AND CONTINUITY.

CHAPTER VII.

THE RATIONAL ALGEBRAIC FUNCTION.

CHAPTER VIII.

CONVERGENCE OF INFINITE SERIES.

CHAPTER IX.

UNIFORM CONVERGENCE OF REAL SERIES.

CHAPTER X.

POWER SERIES.

CHAPTER XI.

OPERATIONS WITH POWER SERIES.

CHAPTER XII.

CONTINUATION OF POWER SERIES.

CHAPTER XIII.

ANALYTIC THEORY OF THE EXPONENTIAL AND LOGARITHM.

CHAPTER XIV.

SINGULAR POINTS OF ANALYTIC FUNCTIONS.

CHAPTER XV.

WEIERSTRASS'S FACTOR-THEOREM.

CHAPTER XVI.

INTEGRATION.

CHAPTER XX.

SIMPLE ALGEBRAIC FUNCTIONS ON RIEMANN SURFACES.

CHAPTER XXI.

ALGEBRAIC FUNCTIONS.

CHAPTER XXII.

CAUCHY'S THEORY AND THE POTENTIAL.

ADDITIONS AND CORRECTIONS.

Page 17, small print, *for* Dane *read* Norwegian.

" 26, Ex. 7, *dele* o or.

" 49, line 19, *after* chief value, *insert*, or chief logarithm.

" 75, line 10, *dele* exactly.

" 85, 86. It is easy to modify the proof of the continuity of a rational integral function so as to make it apply for points other than $x=0$.

" 102. Footnote. English text-books, in general, distinguish between oscillation and divergence of series; Cayley and Stokes however regard oscillating series as divergent and this is the custom of the vast majority of continental mathematicians.

" 225, line 3, *after* But *insert* a value of.

INTRODUCTION TO THE THEORY OF ANALYTIC FUNCTIONS.

CHAPTER I.

THE ORDINAL NUMBER-SYSTEM.

1. Ordinal Numbers. Let us consider a row of objects with regard to their order, say from left to right, freeing our minds from all notions of magnitude. We speak of the first, second, third object, and so on; and by an integer we mean simply a mark which we attach to an object to tell its place in the row. The objects so marked form what may be called the row of natural objects, or the *natural row*.

To count is to label the objects, not primarily to say how many they are; we use this latter notion, but we must emphasize the former. We could begin with the object marked 3 and re-label it 1, then re-label 4 as 2, and so on. This is expressed by writing

$$3-2=1, \quad 4-2=2,$$

meaning that if we begin after the object whose old mark was 2, then the object which was third becomes first, and so on. The beginning after an object instead of with it suggests that our original row might begin after an object; this object we mark 0 and call the origin. If there are objects to the left of the origin we count them in the same way; only we prefix the sign $-$ to show that they are to the left, and call the marks

so altered *negative* integers, distinguishing the old marks as *positive*; the marks now are

$$\ldots -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ \ldots.$$

When one object α is to the left of another α' we say that α comes before α' or is inferior to α', and α' comes after α or is superior to α; and we write $\alpha < \alpha'$, $\alpha' > \alpha$. Here α, α' mean of course integers, positive or negative.

Objects considered as a succession from left to right are *in positive order*; when considered from right to left *in negative order*.

2. **Fractions.** We now attend only to zero,—the object from which (not at which) we begin,—and to the objects on the right of it. We can re-label the alternate objects 2, 4, 6, ..., marking them 1, 2, 3, ...; see fig. 1, where the old names are below the objects, the new above. We must then invent marks for the objects formerly marked 1, 3, 5, ... They can be marked $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$, or $1/2$, $3/2$, $5/2$, ...

0	½	1	3/2	2	5/2	3
x	x	x	x	x	x	x
0	1	2	3	4	5	6

Fig. 1.

Conversely we are at liberty to interpolate alternate objects in a given row 0, 1, 2, 3, ..., only then we mark them $1/2$, $3/2$, $5/2$, and so on.

In the same way we can interpolate two objects between every consecutive two of the given row 0, 1, 2, 3, ..., marking the new objects in order as $1/3$, $2/3$; $4/3$, $5/3$; and so on.

In this way we account for the symbols $1/p$, $2/p$, ..., where p is any positive integer; let us call these symbols positive fractional numbers. By *positive rational* numbers we understand both positive integral and positive fractional numbers.

If we interpolate single objects in the row

$$0,\ 1/2,\ 1,\ 3/2,\ \ldots,$$

we have the same sequence of objects as if we interpolate objects by threes in the row

$$0, \ 1, \ 2, \ldots;$$

and our objects are therefore marked

$$0, \ 1/4, \ 2/4, \ 3/4, \ 1, \ldots.$$

Hence $1/2$ and $2/4$ are two marks for the same object. When we find two marks attached to the same object we say that they are equal; thus we write

$$1/2 = 2/4.$$

A row marked $0, 1/6, 5/6, 1/3, 1/2, 2/3, 1, \ldots$ is to be understood as arising from the interpolation of objects by fives; that is, by introducing the objects $1/6, 2/6, 3/6, 4/6, 5/6, \ldots$, or $1/6, 1/3, 1/2, 2/3, 5/6, \ldots$. As $2/6$ comes before $3/6$ we say that $1/3 < 1/2$.

We may interpolate as many objects as we like in the natural row, and by the principle of the least common denominator we can interpolate so as to explain any assigned positive fractional marks $f_1, f_2, \ldots, f_p$. Also, given any positive rational mark r other than zero, we can interpolate rational marks between 0 and r. When no object can be made to fall between an assigned object and 0, that assigned object must itself be 0.

The same observations apply to the negative numbers.

We can think of an infinity of objects as interpolated in the natural row so that each shall bear a distinct rational number and so that we can assert which of any two comes first. It is to be noticed that as we approach any of the natural objects there is no last fractional mark; that is, whatever object we take there are always others between it and the natural object.

3. Irrational Numbers. Progress in other directions leads us to the notion of square numbers; in considering these

0	1		$3/2$		2					3
x	x	x	x	x	x	x	x	x	x	x
0	1	2	$9/4$	3	4	5	6	7	8	9

Fig. 2.

from the ordinal point of view we re-name our natural row as in fig. 2, where the old names are below, the new above; and we

have to consider how then to bring the omitted objects into the scheme of ordinal numbers. Every object whose new name is fractional had a fractional name, so that the object whose old name was 2 cannot now have a rational name at all. We give it a name which we call irrational; we call it the positive or chief square root of 2 and mark it $\sqrt{2}$ or $2^{\frac{1}{2}}$. As an ordinal number it is perfectly satisfactory, for we know where it comes, whether left or right of any proposed rational number, simply by means of the old naming. Hence it separates *all* the rational numbers into two classes, those on its left and those on its right. A rational number separates all *other* rational numbers into two classes; we put it in one of these classes, and say that it closes that class.

Any process which serves to separate rational numbers into two classes,—those on the left, and those on the right, such that the left-hand class is not closed on the right and the right-hand class is not closed on the left,—leads to the introduction of a new object named by an irrational number.

If a class,—for example the left-hand class,—is closed by an object, we require no new object.

Two numbers rational or irrational,—to fix ideas we will take them both irrational and equal to s and s',—are said to be equal if the rational objects to the left of s are the same as those to the left of s', and the rational objects to the right of s are the same as those to the right of s'. For example, $4^{\frac{1}{4}}$, $2^{\frac{1}{2}}$ effect the same separation of the rational numbers. An equivalent condition for the equality of s, s' is that every rational number to the left of s shall be to the left of s', and every rational number to the left of s' shall be to the left of s.

Between two unequal irrational objects s and s' there must lie rational objects; for since s and s' are not equal there must be a rational number which is before one of the two and not before the other.

It is very important to notice that we have now a closed number-system. When we seek to separate the irrational objects as lying left or right of an object, either the object is rational, or if not it separates rational objects and is irrational; in any case

then it must have for its mark a rational or irrational number, and there is no loophole left for the introduction of new real numbers which separate existing numbers. This is often expressed briefly by saying that the whole system of positive and negative integral, fractional, and irrational numbers is *continuous*, or is *a continuum*.

4. The change of Origin. Let us call the objects associated with the continuous number-system the complete row of objects. This complete row is built into the natural row in the manner described above, the only postulate with regard to the natural row being that of any pair of objects we can say that one lies to the left of the other. In the complete row there are objects *integral as to any object* x in precisely the same way as the integral objects with which we began are integral as to one of them which was taken as zero.

When x is itself an integer this is clear, and has already been used; it is only saying that we can begin our naming from any one of the natural row. But it follows that every object x_1 between 1 and 2 stands to some object x_0, or say x_1', between 0 and 1 in this relation of congruence; that is, if we begin our naming from 1 as origin, then the object x_1 will now have to be called x_1', or, with the notation of § 1,

$$x_1 - 1 = x_1'.$$

		0	x_1'		x_2'	
×	×	×	×	×	×	×
0	x_0	1	x_1	2	x_2	3

Fig. 3.

We can utilize this periodic property to name all objects by integers and *proper* numbers, where by *proper* numbers we mean names of objects between 0 and 1, 0 included and 1 excluded, the last suppositions being made to avoid periphrasis. This is expressed by writing, in the case above,

$$x_1 = 1 + x_1'.$$

So if x_2 lie between 2 and 3 (3 excluded) we have

$$x_2 - 2 = x_0,$$

where x_0 is a proper number. This we can express by

$$x_2 = 2 + x_0,$$

and so on.

In this way the idea of ordinal addition is established in general without reference to quantity.

5. The Decimal System. We can name our complete system methodically as follows. To avoid too many marks we select the 9 digits, 1, 2,..., 9, as marks for objects after zero; and name the following by two marks 10, 11, 12, and so on, up to 19; those following by 20, 21, and so on. In this way all integers are named.

We first interpolate 9 objects between every two consecutive integers; those between 0 and 1 we name ·1, ·2,..., ·9; the objects between 1 and 2 that are congruent to these proper fractions are named 1·1, 1·2,..., 1·9, and so on. The system so gained, together with the integers, is the first decimal system. Between every two consecutive objects of this system we interpolate 9 new objects, and we name those between 0 and ·1 ·01, ·02,..., ·09; those between ·1 and ·2 as ·11, ·12,..., ·19, and so on. We thus obtain a system which with the first decimal system is the second decimal system. We can proceed as far as we please in this way. And we can freely change the origin to any object so gained, by the addition or subtraction of decimals, performed of course in the usual way, but interpreted solely with reference to before and after, not to how much. The pth decimal system is, whatever p may be, only a part of the rational objects, and includes no irrational object. But by choosing p large enough we can separate any two rational objects by this pth system. The proof is as follows:—

(1) Let the two rational objects be 0, r where r is positive. We can take p sufficiently large to make the first proper number other than zero in the pth decimal system lie between 0 and r; since if r have the name p'/p'' it coincides with or comes after $1/p''$, and therefore we have only to interpolate more than p'' objects between 0 and 1 to get an object nearer to 0 than either of $1/p''$, r is.

(2) And secondly, if the two rational objects be r, r' and r, r' have a common denominator l, we have only to take more than l objects to separate r and r'.

Thus our decimal system allows us to separate any two rational objects by taking p large enough.

We can, for example, separate 0 and any positive rational object by the object $1/10^p$ by taking p large enough. This is expressed by saying that the *limit* of $1/10^p$ when p increases, or when p is infinite, is zero.

6. The Infinite Decimal. Let us consider how the object $1/3$ can be named in the decimal system. We have

$$\begin{aligned} 1/3 &= {\cdot}3 + 1/30 \\ &= {\cdot}33 + 1/300 \\ &= {\cdot}333 + 1/3000, \end{aligned}$$

and generally

$$1/3 = {\cdot}333\ldots\text{to } p \text{ places} + 1/3{\cdot}10^p.$$

Whatever rational object we may select on the left of $1/3$, we can by taking p large enough assign a decimal, ${\cdot}333\ldots$to p places which will separate r and $1/3$.

We express this by saying that the *limit* of ${\cdot}3$, ${\cdot}33$, ${\cdot}333$, ... ${\cdot}333\ldots$to p places is $1/3$ when p is infinite.

We have here a sequence of decimals, namely ${\cdot}3$, ${\cdot}33$, ${\cdot}333$, and so on, which can be carried as far as we please and of which we know the digit in any place. We speak of such a sequence as an *infinite decimal.*

It is known from ordinary arithmetic that, as in the above instance, from any rational object r arises an infinite decimal (that is, a sequence of decimals) or else a terminating decimal; in this latter case the sequence is ultimately a mere repetition of the rational object itself; it is easy to modify the proof in the case of $1/3$ so as to show that the sequence will separate r from any rational object on its left; and we again define the object r as the *limit* of the sequence. In the case of the terminating or ordinary decimal the limit is attained.

We can, then, name any rational object by a decimal,

terminating or infinite. As a matter of fact the decimal when infinite must sooner or later be periodic; but this is not material to our purpose.

Now suppose an infinite decimal given; let it be

$$0.\alpha_1\,\alpha_2\ldots\alpha_p\ldots,$$

where α_p is a digit which one is in a position to assign when p is assigned, but the number of digits is unlimited[1].

Expressing any proposed rational number as a decimal, we can compare the two decimals; either they are the same, i.e. they have the same digits in the pth place for every p, or we can assign their order. Thus any infinite decimal either leads to a rational object or separates rational objects into two classes. By the definition of an irrational number the infinite decimal leads, in the latter case, to an irrational number d. Hence every infinite decimal leads to a number, rational or irrational. Since the infinite decimal which leads to an irrational number d separates d from any rational object to the left of d, we say as before that d is the *limit* of the infinite decimal. It is convenient to express both the infinite decimal and its limit in the same way; e.g. to say that ·333... is 1/3.

The infinite decimal is a sequence of objects in positive order. As a type of a sequence in negative order we may consider a negative decimal, for example,

$$-\cdot 6,\ -\cdot 66,\ -\cdot 666,\ldots$$

which will in the same way define a negative object, easily identified with $-2/3$. Another sequence in negative order is obtained by adding 1 to each object of the above sequence, that is, by taking the sequence

$$\cdot 4,\ \cdot 34,\ \cdot 334,\ \cdot 3334,\ldots,$$

the object defined here is $1-2/3$ or $1/3$.

Thus 1/3 can be defined either by a sequence in positive order or by a sequence in negative order.

Generally let d be the limit of an infinite decimal

$$d_1,\ d_2,\ d_3,\ldots,\ d_p,\ \ldots,$$

[1] We exclude, in order to avoid two decimal names for one object, the case where every digit after an assigned one is 9.

where $$d_p \text{ is } 0 . \alpha_1 \alpha_2 \ldots \alpha_p;$$
then d is an object between 0 and 1 (0 included); and
$$(d_1 + 1/10), (d_2 + 1/10^2), \ldots, (d_p + 1/10^p), \ldots$$
is a sequence in negative order, which is not in the form of a decimal sequence; but
$$1 - (d_1 + 1/10), 1 - (d_2 + 1/10^2), \ldots, 1 - (d_p + 1/10^p), \ldots$$
is a decimal sequence and leads to an object; hence it is clear that the sequence
$$d_1 + 1/10, d_2 + 1/10^2, \ldots, d_p + 1/10^p, \ldots$$
does also name an object. To this object we can assign an infinite decimal d'. It is important to prove that d' is none other than d itself. If not, it is clear that d' is to the right of d; that is, $d < d'$. We have then for every p the following ordering
$$d'_p < d < d' < d_p + 1/10^p;$$
and $$d_p + 1/10^p < d + 1/10^p;$$
whence we have the following ordering
$$d < d' < d + 1/10^p,$$
whatever integer p may be.

But d and $d + 1/10^p$ are the same to as many decimal places as we please, whereas d and d', being supposed distinct, separate at an assignable decimal place. Hence d' is not distinguishable from d; there are not two objects, but one object.

For example, the two sequences
$$\cdot 4, \cdot 41, \cdot 414, \cdot 4142, \ldots$$
and $$\cdot 5, \cdot 42, \cdot 415, \cdot 4143, \ldots$$
define one and the same object, named $2^{\frac{1}{2}} - 1$.

It is clear that the same reasoning applies to other than proper numbers.

7. Distance, Point, and Angle. The distance from one point to another can be measured in terms of a unit length by seeing first how many units there are, then how many tenths in the part left over, and so on. We thus obtain a decimal, terminating

or not, which defines a number. By the distance we understand this number.

Conversely we *postulate* that every positive decimal d, infinite or not, that is every distance, can be represented by a terminated line $\mathrm{o}d$, or simply by the point d itself, d being on the right of o. Let d' be any other point on the right of o; if d' is on the right of d we say the distance d' is greater than the distance d, and d less than d', and write

$$d' > d, \quad d < d',$$

in a new sense.

The decimal sequence $d_1, d_2, d_3, \ldots, d_p, \ldots$, is an ascending sequence of distances defining by the postulate of this article the point d. Similarly d can be defined by

$$d_1 + 1/10, d_2 + 1/10^2, \ldots, d_p + 1/10^p, \ldots.$$

A similar postulate is required for angles. We have a natural measure of amount of turn, namely, the circumference of a unit circle. This unit can be divided geometrically into 2^p equal parts. Out of sequences formed from these we have to define the number $1/2\pi$, and then we postulate the existence of a point whose distance along the arc is this number. With the aid of infinite series the problem can be simplified.

What has been said of the theory of ordinal numbers is necessary for the subsequent analytic theory, and is, we hope, sufficient to render the argument intelligible. As our subject is functions of complex numbers, we shall tacitly assume further results concerning real numbers, as, for example, that we can effect the elementary operations with irrational numbers; and that nothing new results.

For a detailed account of real numbers we may refer to Fine's *Number-system of Algebra*.

CHAPTER II.

THE GEOMETRIC REPRESENTATION OF COMPLEX NUMBERS.

8. **Introductory remarks.** We denote by i one of the roots of the equation

$$x^2 + 1 = 0.$$

The expression $\xi + i\eta$, where ξ and η are any real numbers, is a *complex number*. When $\xi = 0$ the number is called *imaginary*. Thus i itself is a number; we call it the *imaginary unit*.

We assume that the elementary operations of addition, subtraction, multiplication and division have been defined and justified for complex numbers. What is said in elementary algebra on complex numbers covers our assumption, when we admit,—as we have done,—that we can effect these same operations on irrational numbers. We must notice that *the totality of numbers $\xi + i\eta$ is a closed system with respect to the elementary operations*; that is, the result of an operation is again a number, real or complex.

The object of this book is to discuss the properties of such expressions as grow naturally from the repetition of elementary operations, when applied to complex numbers. An adequate geometric representation of a complex number is almost indispensable, and we shall consider in this chapter the representation by means of directed distances, with some immediate consequences.

But in using geometric intuitions, and thus beginning what is called the geometric theory of functions of a complex variable, we must emphasize one lesson of experience: that the intuitional

method is not in itself sufficient for the superstructure. It has been found that only by the notion of number, somewhat as sketched in the first chapter, can fundamental problems be solved. If however we are prepared to replace, when occasion arises, these geometric intuitions,—which are point, distance, and angle,—by numbers, as in § 7, then, and only then, is the use of geometry thoroughly available.

This implies no reflexion on the customary applications of geometry to algebra, or conversely, which are made in analytic geometry or elementary calculus. The point is that elementary algebra, as commonly presented, is imperfect and is itself to some extent intuitive. It is at this stage that it becomes necessary to insist on some refinements in our notions of algebra, and we may reasonably attempt a parallel precision in our notions of geometry. The distance and the number must always correspond, and in so far as we can know and say what we mean by the one we must be prepared to know and say what we mean by the other.

9. The Axis of Real Numbers. As already said in § 7, we represent real numbers by distances or *strokes* measured from an origin along a straight line, which we call the axis of real numbers, a stroke to represent 1, or unit stroke, being arbitrarily selected. The unit stroke will always be taken to the right, or eastward. Then all positive strokes are eastward, all negative ones westward. The point which ends the stroke bears the same name or real number as the stroke itself. This way of naming points, being already used in analytic geometry,—singly for the line, doubly for the plane, and triply for space,—need not be enlarged upon.

We need only remark that in the addition of real strokes, the second is to begin where the first ends, and so on.

We next consider a possible interpretation of the sign $-$.

The numbers n, $-n$ are represented by strokes of equal lengths, but opposite directions. Now if the stroke n be regarded as a material arrow with a fixed end at 0, the direction of the arrow cannot be reversed without moving it out of the line

The simplest method for securing this reversal of direction is to rotate the arrow, in a plane that passes through it, either positively or negatively, through two right angles. This affords the following interpretation for the sign $-$.

The sign $-$ attached to a positive stroke can be treated as an operator which rotates the stroke about its fixed extremity O *through two right angles; that is, the sign $-$ gives a half-turn to the stroke.*

From this point of view the statement

$$(-)\times(-) \text{ is } +$$

means that two half-turns bring the arrow back to its primitive position.

Such an equation as

$$(-2)\times 3 = -6$$

is capable of the following two interpretations:—

(*a*) the product of the two numbers -2, 3 is the number -6;

(*b*) the result of doubling the stroke 3 and reversing its direction is the negative stroke -6.

In (*a*) -2 is a number, but in (*b*) it is an operator which means 'double and reverse.'

10. Imaginary Numbers, and the Axis of Imaginary Numbers. The real number x is equal to $x\times 1$. Hence x can be regarded as an operator. It is then a product of two independent operators, the one changing the size of the unit stroke and equivalent to a *stretch*, the other changing the sign and equivalent to a *turn*; when x is positive only the former comes into play.

While the amount of stretching may be made what we please by a suitable choice of x, the amount of turning is restricted to an odd or even number of half-turns. Is it possible to remove this restriction and allow the unit stroke to be turned as freely as it can be stretched, by allowing x to take imaginary or complex values? We shall show in § 15 that this question can be answered unreservedly in the affirmative. Our immediate task is to find an interpretation for purely imaginary numbers as operators.

Let the plane of the paper be the plane in which the rotations take place, and let the axis of real numbers go from west to east, as in fig. 4. Let the order east, north, west, south denote the positive order of turning.

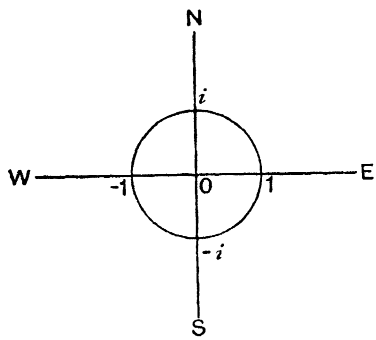


Fig. 4.

The operator which turns the unit stroke positively through a right angle we call i, without assuming any connexion between i and $\sqrt{-1}$. Hence $i.1$, or briefly i, is a directed line of unit length which begins at O and ends at the point marked i in fig. 4; such a line is called a *stroke* equally with lines along OE or OW. The operator i, when applied to the stroke i, turns that stroke positively through a right angle and produces the stroke from O to -1. Thus

$$i\,.\,i\,.\,1 = (-1) \times 1,$$

giving $\qquad i^2 = -1,\ i\,.\,1 = \sqrt{-1}$; so that $i = \sqrt{-1}$.

This is the geometric representation of i (§ 8); *the imaginary unit is represented, as we have seen, by a stroke of unit length along an axis perpendicular to the axis of real numbers and called the axis of imaginary numbers.*

All the imaginary numbers are built up from the imaginary unit i in the same way as the real numbers are built up from 1. The product and ratio of two imaginary numbers are real, just as the product and ratio of two negative numbers are positive.

The numbers mi, where m is real, can be regarded as naming either the points at the ends, other than O, of the corresponding strokes; or the strokes themselves. Further, such strokes as $3i$,

$-3i$, can be derived from the unit stroke by a stretch 3 followed by positive turns through one, three right angles respectively.

11. Strokes. Let us now understand generally by a *stroke* a straight line of definite length and definite direction which lies in an assigned plane. This definition includes as particular cases the directed length that has been used in the preceding paragraphs.

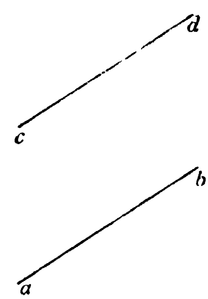


Fig. 5.

Two strokes $\overline{ab}$, $\overline{cd}$ (fig. 5) are said to be equal when they are of equal lengths and are drawn along parallel lines in the same sense. Observe that $\overline{ab}$ is not equal to $\overline{dc}$, for $\overline{dc} = -\overline{cd}$.

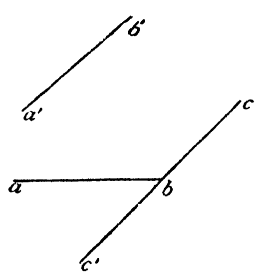


Fig. 6.

Addition. The sum of two strokes $\overline{ab}$, $\overline{a'b'}$ is defined in the following manner. Draw the stroke $\overline{bc}$ equal to $\overline{a'b'}$; then the sum of $\overline{ab}$, $\overline{a'b'}$ is the same as the sum of $\overline{ab}$, $\overline{bc}$, and this latter sum is defined to be $\overline{ac}$. Thus the equation

$$\overline{ab} + \overline{bc} = \overline{ac}$$

holds even when $\overline{ab}$, $\overline{bc}$ are not in one and the same straight line.

More generally the equation

$$\overline{ab} + \overline{bc} + \overline{cd} + \overline{de} + \overline{ef} + \overline{fg} = \overline{ag}$$

holds not only when $\overline{ab}$, $\overline{bc}$, etc., are strokes along the axis of real numbers, but also when the directions of the component strokes are any whatsoever.

Subtraction. Such a difference as $\overline{ab} - \overline{a'b'}$ is defined as the sum of the strokes $\overline{ab}$, $\overline{b'a'}$ and is equal to $\overline{ac'}$ (fig. 6) where $\overline{bc'} = -\overline{bc}$.

It is convenient in comparing strokes to draw them from a common origin o. Then each stroke $\overline{op}$ is determined by its terminal point p, and there is a 1, 1 correspondence between the strokes $\overline{op}$ and the points p. The strokes may therefore be characterized by their terminal points p. We shall show in the next paragraph that we can establish a 1, 1 correspondence between the points p of the plane and the numbers in the complete number-system.

The following are convenient constructions for $\overline{op} + \overline{oq}$ and

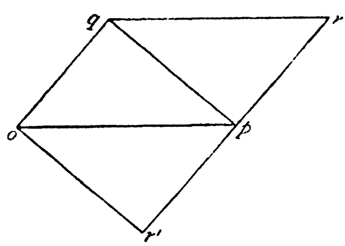


Fig. 7.

$\overline{op} - \overline{oq}$. Complete the parallelogram $oprq$ whose sides are $\overline{op}$, $\overline{oq}$, and produce rp to r' so that $r'p$, pr are equal in length. Then

$$\overline{op} + \overline{oq} = \overline{op} + \overline{pr} = \overline{or}.$$

$$\overline{op} - \overline{oq} = \overline{op} + \overline{qo} = \overline{op} + \overline{pr'} = \overline{or'} = \overline{qp}.$$

NOTE. The rule for the addition of strokes applies to velocities, momenta, accelerations, etc.

12. Complex Numbers and the Points of a Plane. Let ξ, η be real numbers; the complex number $x = \xi + i\eta$ is the

sum of ξ and $i\eta$, which can be represented by strokes at right angles of length ξ, η as in fig. 8. The sum of these two strokes

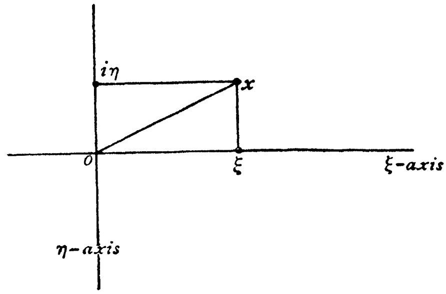


Fig. 8.

is $\overline{ox}$. Hence the number x can be associated with the stroke $\overline{ox}$, or the point x. That is, the point (or the stroke from the origin to the point) represents the number, and the number names the point (or stroke). Here then we have a starting-point for the geometric treatment of complex numbers. The points on the straight lines which are labelled ξ-axis and η-axis (fig. 8) represent graphically all real and all imaginary numbers; and complex numbers $\xi + i\eta$ are visualized by points x with rectangular coordinates (ξ, η).

This method of representing complex numbers was explained in a work by Argand (in 1806). Gauss had discovered it probably as early as 1799, but published nothing on the subject till 1831. A Dane, Wessel, appears to have first published the method in 1797.

13. Absolute Value and Amplitude of $x = \xi + i\eta$. Let the polar coordinates of (ξ, η) be ρ, θ; the positive length ρ is called the *absolute value* of x and is denoted by $|x|$.

The angle θ, and equally so the angles $\theta + 2n\pi$ where n is any positive or negative integer, determine the direction of the stroke x. Any one of these angles may be called the *amplitude* of the stroke. Frequently it is convenient to select a definite one of the set, θ_0; we shall do so by supposing

$$-\pi < \theta_0 \leqq \pi;$$

we shall call this *the chief amplitude*, and denote it when convenient by $\mathrm{Am}\, x$, while denoting any amplitude by $\mathrm{am}\, x$.

The many-valuedness of an amplitude deserves careful attention in this subject, as the explanation of many-valuedness in general depends on it.

The following relations connect ξ, η, ρ, θ :—

$$\xi = \rho \cos \theta,\ \eta = \rho \sin \theta,\ \rho = +\sqrt{\xi^2 + \eta^2},\ \tan \theta = \eta/\xi\,;$$

hence $$x = \xi + i\eta = \rho(\cos \theta + i \sin \theta),$$

or, as we may write it,

$$x = \xi + i\eta = \rho \operatorname{cis} \theta.$$

Later on, when we have defined the symbol e^x, where x is complex, we shall find that $$\xi + i\eta = \rho e^{i\theta}.$$

Interpretation of the symbol $\rho \operatorname{cis} \theta$ as an operator. Each factor of the expression $\rho \operatorname{cis} \theta$ conveys important information. The stroke x can be derived from the stroke 1 by a stretch which changes the length 1 to the length ρ and by a turn of the resulting stroke through an angle θ; hence the following possible interpretation for the equation

$$\rho \operatorname{cis} \theta \,.\, 1 = \rho \operatorname{cis} \theta :—$$

The expression $\rho \operatorname{cis} \theta$ can be regarded as an operator which stretches the stroke 1 into a stroke ρ and then turns this stroke ρ through an angle θ. This interpretation includes the earlier interpretation of imaginary and negative numbers, as is seen at once by putting $\theta = \pi/2,\ \pi,\ 3\pi/2$.

14. Addition of Two Complex Numbers. In algebra the sum of the two complex numbers $\xi_1 + i\eta_1$, $\xi_2 + i\eta_2$ is $\xi_1 + \xi_2 + i(\eta_1 + \eta_2)$. Thus addition connects the points (ξ_1, η_1), (ξ_2, η_2) with the point $(\xi_1 + \xi_2, \eta_1 + \eta_2)$. But there is no difficulty in proving geometrically that this third point is a corner of the

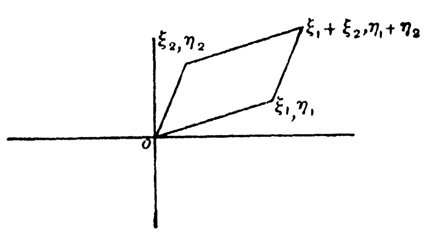


Fig. 9.

parallelogram constructed on the lines from o to (ξ_1, η_1) and (ξ_2, η_2); *hence the result obtained by algebraic addition of two numbers agrees with the result obtained by the addition of the corresponding strokes according to the parallelogram law.*

It is evident from fig. 9 that *the absolute value of the sum of two complex numbers x_1, x_2 is, in general, less than the sum of the absolute values of these numbers.* For $|x_1|$, $|x_2|$ are the lengths of the sides of the parallelogram and the sum of these two sides is greater than the diagonal whose length is $|x_1+x_2|$. In the special case where x_1, x_2 have the same amplitude we have

$$|x_1+x_2| = |x_1| + |x_2|.$$

More generally we may say that *the absolute value of the sum of several complex numbers cannot be greater than the sum of the absolute values of the component numbers.* This is merely the analytic statement of the geometric theorem that the length of one side of a closed polygon is less than the sum of the lengths of the other sides, the words *less than* being replaced by *equal to* when all the other sides have the same amplitude.

Subtraction. Since the subtraction of strokes is merely the addition of other strokes and the subtraction of complex numbers is the addition of other complex numbers, the subtraction of two numbers is interpreted geometrically by the rule given for the subtraction of two strokes.

It is well to observe that the stroke from the head of x_1 to

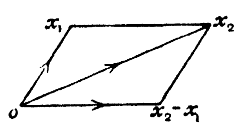


Fig. 10.

the head of x_2 is the stroke $x_2 - x_1$ (fig. 10). It is often spoken of as the *change* or *increment* of a variable stroke x as it passes from the value x_1 to the value x_2. The notation $x_2 - x_1$ replaces the temporary notation $\overline{x_1x_2}$.

15. Ratio and Multiplication. Let a, b, be two complex numbers; then b/a in the equation

$$(b/a) \times a = b$$

can be interpreted either as a number, or as an operator. From the latter point of view it changes the stroke a into the stroke b (1) by stretching the stroke of length $|a|$ into one of length $|b|$, (2) by turning the resulting stroke until its amplitude becomes equal to that of b. Hence the operator b/a stretches by an amount $\dfrac{|b|}{|a|}$ and turns by an amount $\text{am}\, b - \text{am}\, a$.

Just as $\rho \operatorname{cis} \theta$ has a dual signification (1) as an operator, (2) as a stroke, so b/a can be interpreted not merely as an operator but also as a stroke of length $\dfrac{|b|}{|a|}$ and direction $\text{am}\, b - \text{am}\, a$. This stroke b/a is generated from the stroke 1 by the operator b/a.

We have then a geometric interpretation for b/a as a stroke; and we see that *the absolute value of b/a is the ratio of the absolute values of b and a, and the amplitude of b/a is the change of amplitude in passing from a to b.*

That is, $\left|\dfrac{b}{a}\right| = \dfrac{|b|}{|a|}$ and $\text{am}\left(\dfrac{b}{a}\right) = \text{am}\, b - \text{am}\, a$.

But we cannot always say, for chief amplitudes, that $\text{Am}\left(\dfrac{b}{a}\right) = \text{Am}\, b - \text{Am}\, a$; for example, if $\text{Am}\, b = \pi$, $\text{Am}\, a = -\pi/2$, then $\text{Am}\left(\dfrac{b}{a}\right) = -\pi/2$.

It is easy to construct geometrically the quotient b/a. In fig. 11, the two triangles

$$0,\ a,\ b, \text{ and } 0,\ 1,\ b/a,$$

are directly similar; for the angles at 0 are equal to $\text{am}\, b - \text{am}\, a$ and the sides about these angles are proportional.

Ex. 1. *Calculate the absolute value and amplitude of* $\dfrac{2+i}{3-i}$.

Ex. 2. *Prove geometrically that* $\left|\dfrac{\alpha + i\beta}{\beta + i\alpha}\right| = 1$.

We observe that if $b/a = c$, then $b = ac$; hence fig. 11 provides us with a geometric construction for the product of a and c.

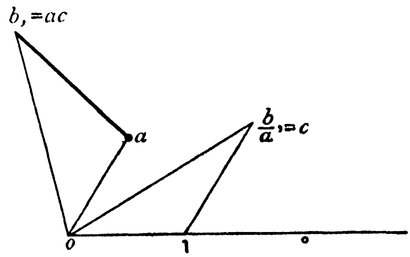


Fig. 11.

The figure shows that

$$|ac| = |a|\,|c| \text{ and } \operatorname{am}(ac) = \operatorname{am} a + \operatorname{am} c.$$

Thus the symbol a in ac can be interpreted as an operator that stretches the length $|c|$ of c in the ratio $|a| : 1$, and increases the amplitude of c by the angle am a. In words:—

The absolute value of a product ac is the product of the absolute values of a and c, while the amplitude of ac is the sum of the amplitudes of a and c.

The same rules follow also from the direct division and multiplication of $\rho_1 \operatorname{cis} \theta_1$ and $\rho_2 \operatorname{cis} \theta_2$. For example,

$$\begin{aligned}\rho_1 \operatorname{cis} \theta_1 \times \rho_2 \operatorname{cis} \theta_2 &= \rho_1\rho_2 (\cos\theta_1 + i \sin\theta_1)(\cos\theta_2 + i\sin\theta_2)\\ &= \rho_1\rho_2 \{(\cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2)\\ &\qquad + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)\}\\ &= \rho_1\rho_2\{\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)\}\\ &= \rho_1\rho_2 \operatorname{cis}(\theta_1+\theta_2).\end{aligned}$$

We have assumed the addition-theorems of the sine and cosine; but it is worthy of note that the matter can be so presented as to prove these fundamental theorems. For since $\operatorname{cis}\theta$ is a turn through θ, the meaning of $\operatorname{cis}\theta_1 \operatorname{cis}\theta_2$ is a turn first through θ_2 and then through θ_1. It is therefore a turn through $\theta_1 + \theta_2$, or is $\operatorname{cis}(\theta_1 + \theta_2)$. Hence by equating real and imaginary parts we have the addition-theorems. The theorem

$$\operatorname{cis}\theta_1 \operatorname{cis}\theta_2 \ldots \operatorname{cis}\theta_n = \operatorname{cis}(\theta_1 + \theta_2 + \ldots + \theta_n),$$

which is known in connexion with Demoivre's theorem, states

that n separate turns $\theta_n, \theta_{n-1}, \ldots, \theta_2, \theta_1$ amount on the whole to a single turn $\theta_1 + \theta_2 + \ldots + \theta_n$.

Ex. If $2\cos\theta = a + 1/a$, prove that $2\cos n\theta = a^n + 1/a^n$, where n is any integer.

16. The nth Roots of Unity. An important special case of the theorem

$$\operatorname{cis}\theta_1 \operatorname{cis}\theta_2 \ldots\ldots \operatorname{cis}\theta_n = \operatorname{cis}(\theta_1 + \theta_2 + \ldots + \theta_n)$$

is that which arises when $\theta_1, \theta_2, \ldots, \theta_n$ are all put equal to $2m\pi/n$, where m is an integer. The theorem becomes

$$(\operatorname{cis} 2m\pi/n)^n = \operatorname{cis} 2m\pi = 1.$$

Hence $\operatorname{cis} 2m\pi/n$ is an nth root of unity. By making

$$m = 0, 1, 2, \ldots, n-1,$$

we obtain n distinct roots, and no additional distinct roots are obtainable by giving m other integral values.

Geometrically $\operatorname{cis} 2m\pi/n$ is a turn through m/n of four right angles; and the continuous repetition of the single turn $2\pi/n$ will give all the nth roots of unity. The n points which represent the roots form the vertices of a regular polygon. The figure shows this for $n = 6$.

Every equation $x^n - 1 = 0$, where $n = 3, 4, 5, \ldots$, has certain roots, called *primitive* nth roots, which satisfy no equation of similar form and lower degree. For example, $x^6 = 1$ is satisfied by two primitive roots of this kind, namely, $\operatorname{cis} 2\pi/6$ and $\operatorname{cis} - 2\pi/6$; the other roots satisfy one or more of the equations $x - 1 = 0$, $x^2 - 1 = 0$, or $x^3 - 1 = 0$, and are not primitive. The

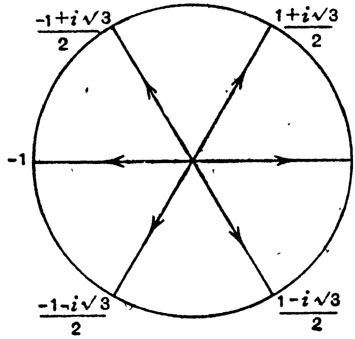


Fig. 12.

primitive nth roots of unity are, then, those turns which, with their repetitions, give all the nth roots of unity.

The case $n=3$ deserves notice. The cube roots of 1 are 1, $\cos 2\pi/3 \pm i \sin 2\pi/3$; that is,

$$1, \frac{-1+i\sqrt{3}}{2}, \frac{-1-i\sqrt{3}}{2}.$$

For the second root we shall use on occasion the Greek letter υ; thus υ denotes a turn through $2\pi/3$, or the complex number $-\frac{1}{2}+i\frac{\sqrt{3}}{2}$.

Ex. 1. Verify, by a diagram, that the cube roots of unity are $1, \upsilon, \upsilon^2$, and that their sum is zero.

Ex. 2. Draw the strokes which represent the square roots of i and write down these roots in the form $\xi+i\eta$.

Ex. 3. Find the fifth roots of unity. Prove that the sum of the nth powers of these fifth roots is 5 or 0, according as the positive integer n is or is not a multiple of 5.

Ex. 4. Given that $\epsilon = \text{cis}\, 2\pi/5$, prove that $(\epsilon^2-\epsilon^3)(\epsilon^4-\epsilon)=5^{1/2}$.

17. The nth Power and nth Root of a Stroke. The stroke a^n is constructed by the rule for a product. We make the triangles $0, 1, a$; $0, a, a^2$; etc. all similar; then the last point of the nth triangle belongs to the stroke a^n.

If we continue the series of triangles in the reverse direction, we get the negative powers $1/a$ or a^{-1}, $1/a^2$ or a^{-2}, etc.

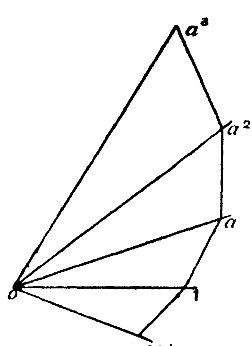


Fig. 13.

If $|a| = 1$, the points all lie on a circle of radius 1. The curve on which the points lie when $|a|$ is not 1 is an equiangular spiral. See § 29.

Let b be an nth root of a, so that $b^n = a$. Then the length of b is the positive nth root of the length of a, and the amplitude of b is an nth part of an amplitude of a. If θ be one amplitude of a, all the amplitudes of a are given by $\theta + 2m\pi$. These amplitudes will have n *different* nth parts, namely,

$$\theta/n,\ \theta/n + 2\pi/n,\ \theta/n + 4\pi/n, \ldots,\ \theta/n + 2(n-1)\pi/n;$$

all the rest are congruent to these n with respect to 2π. Thus we have as in the last paragraph n distinct directions for b, and therefore n nth roots of a. Clearly the n corresponding points are the corners of a regular polygon; also it is evident that the remaining $n-1$ nth roots can be derived from an assigned nth root by multiplying it by the nth roots of unity.

We shall define the *chief nth root* of a as the one whose amplitude is θ/n, where $-\pi < \theta \leqq \pi$. This root will be denoted by $a^{1/n}$, any root by $\sqrt[n]{a}$.

18. To find the point which divides in a given ratio r the stroke from a_1 to a_2.

Let x be the point. What is meant is this: when a point moves from a_1 to a_2 by way of x the stroke $a_2 - a_1$ is resolved into two strokes $x - a_1$, $a_2 - x$. And we may say that x divides the stroke from a_1 to a_2 (or the points a_1, a_2 in the order named) in the ratio $(x - a_1)/(a_2 - x)$. We have then

$$x - a_1 = r(a_2 - x),$$

or

$$x = \frac{a_1 + ra_2}{1 + r}.$$

To understand clearly what is implied by this formula when r is not real, consider this question: opposite corners of a square are a_1 and a_2, what names are to be assigned to the other two corners?

Here we are to express x and y in terms of a_1 and a_2 (fig. 14).

Now $a_1 - x$ is obtained by turning $x - a_2$ through a positive right angle. Hence

$$a_1 - x = i(x - a_2).$$

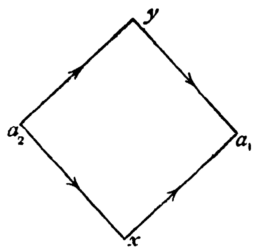


Fig. 14.

Similarly $$a_1 - y = -i(y - a_2).$$

Here r is i for the point x and $-i$ for y.

Ex. 1. Given two opposite points of a regular hexad (namely, the corners a_1, a_4 of a regular hexagon) express the other four points in terms of a_1 and a_4.

Ex. 2. Mark the points which divide 0, 1 in the ratios $\pm(1+i)$.

19. The Centroid of a System of Points. If we let $r = 1$, then we have for the middle point of the join of a_1, a_2, or briefly the middle point of a_1, a_2, the expression $\frac{a_1 + a_2}{2}$. In general we define the centroid of m points $a_1, a_2, \ldots, a_m$ as the point

$$\frac{a_1 + a_2 + \ldots + a_m}{m},$$

or $\sum_{\lambda=1}^{m} a_\lambda / m$; so that in particular the centroid of a_1, a_2 is the middle point of the stroke from a_1 to a_2.

Let $\bar{a}$, $\bar{b}$ be the centroids of the sets $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_n$; the centroid of the $m+n$ points is

$$\frac{\Sigma a_\lambda + \Sigma b_\lambda}{m+n}, = \frac{m\bar{a} + n\bar{b}}{m+n},$$

and divides therefore the centroids of the two systems of m and n points in the ratio n/m.

EXAMPLES.

1. The two triads a, b, c; x, y, z; form similar triangles if

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ x & y & z \end{vmatrix} = 0.$$

2. If the points x, y, z divide the strokes $c-b$, $a-c$, $b-a$ in the same ratio r, and the triangles x, y, z and a, b, c are similar, either $r=1$ or both triangles are equilateral.

3. Let a, b, c, d form a parallelogram of which a, d are opposite points. Let a, b, x; c, d, y; a, d, z be similar. Prove that x, y, z is similar to each of them.

4. If a, x, y; y, b, x; x, y, c are similar, each is similar to a, b, c.

5. If $x^2+y^2=1$, prove that x, y are ends of conjugate radii of an ellipse whose foci are ± 1.

6. Equilateral triangles are described on the sides of a given triangle, all outwards or all inwards. Prove that their centroids form an equilateral triangle.

7. Prove that $\lambda+\mu v+\nu v^2$, where λ, μ, ν are integers whose sum is 0 or ± 1, represents the points of a quilt formed by regular hexagons.

CHAPTER III.

THE BILINEAR TRANSFORMATION.

20. The One-to-one Correspondence. An equation in x and y will be used to establish a correspondence between these variables. It is often convenient to represent the points x and y *in different planes*; in each of these planes the origin and the point 1 are selected arbitrarily, except that the unit of length is supposed usually to be the same for both planes. The equation between x and y establishes, then, a correspondence between the points of the two planes, and either plane is said to be *mapped* on the other. One important way of forming a mental image of the mapping is to draw a series of paths in the one plane (for example straight lines parallel to an axis, or concentric circles) and to determine the corresponding paths in the other plane; a y-path which arises in this way from an x-path is said to be *the map of the x-path.*

When the equation is

$$y=(ax+b)/(cx+d) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1),$$

the correspondence of the two planes is 1, 1; that is, one and only one y corresponds to every x, and conversely. There is one possible exception in the finite part of the x-plane, namely, the point $x=-d/c$; to remove this exception we treat $y=\infty$ as a single point and say that there is in the y-plane *one and only one point at* ∞. With this convention $x=\infty$ represents a point; to $x=\infty$ corresponds one and only one point in the y-plane, namely, a/c. And now the correspondence is 1, 1 whatever value

finite or infinite, be given to x (or y). The equation (1) is the most general algebraic equation giving a 1, 1 correspondence between the planes of x and y; hence the importance of the transformation (1), or the *bilinear transformation* as it is called.

It must be noticed that ∞ is a point only by virtue of a very convenient agreement. It is an artificial point. So we can regard the artificial value or number ∞ as attached to the point ∞.

21. Inverse Points. The bilinear transformation depends only on the ratios of the given constants a, b, c, d; and is therefore given when three pairs of values of x and y are given.

Let x_1, y_1 and x_2, y_2 be corresponding values; we have

$$y - y_1 = \frac{ax+b}{cx+d} - \frac{ax_1+b}{cx_1+d} = \frac{(ad-bc)(x-x_1)}{(cx+d)(cx_1+d)};$$

similarly
$$y - y_2 = \frac{(ad-bc)(x-x_2)}{(cx+d)(cx_2+d)}.$$

Therefore
$$\frac{y-y_1}{y-y_2} = \frac{cx_2+d}{cx_1+d} \cdot \frac{x-x_1}{x-x_2} = k\frac{x-x_1}{x-x_2} \quad \text{............ (2)},$$

where k depends on x_1, x_2, but not on x.

Suppose that x varies subject to one or other of the conditions

$$\left|\frac{x-x_1}{x-x_2}\right| = \text{constant}, \quad \text{am}\left(\frac{x-x_1}{x-x_2}\right) = \text{constant},$$

and that x_1, x_2 are arbitrary but fixed points of the x-plane; what curves are traced out by x?

This question must be answered before the full significance of equation (2) can be appreciated.

I. Let the amplitude of $(x-x_1)/(x-x_2)$ be constant and

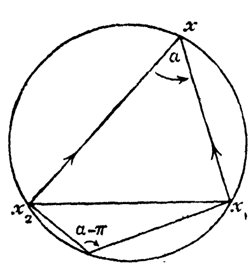


Fig. 15.

equal to α. Then the angle $x_2 x x_1$ is α, so that x can take all positions on a certain circular arc whose ends are x_1, x_2. Observe that for the complementary arc which makes up the complete circle, the amplitude of our ratio, though still constant, is $\alpha - \pi$. Hence we say:—

When the amplitude of $(x-x_1)/(x-x_2)$ *is given,* x *moves on an arc of a circle.*

II. Next let the absolute value of the ratio be constant and equal to ρ, say. It can be proved that x moves on a circle, and that the circle is in this case complete.

This proposition was not given by Euclid, and its fundamental character is insufficiently recognized by many of the modern text-books on elementary geometry. One proof is as follows:—

Choose the point a (fig. 16) on the join of x_1 and x_2 so as to

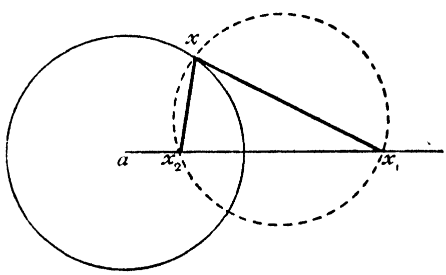


Fig. 16.

make the angle $x_2 x a$ equal to the angle $a x_1 x$. Then the triangles $x_2 x a$ and $x x_1 a$ are equiangular; and therefore the sides of the one are proportional to the sides of the other, giving

$$\left|\frac{x-x_1}{x-x_2}\right| = \left|\frac{a-x}{a-x_2}\right| = \left|\frac{a-x_1}{a-x}\right|$$

The first of these numbers is ρ; therefore

$$\left|\frac{a-x}{a-x_2}\right| \times \left|\frac{a-x_1}{a-x}\right| = \rho^2,$$

that is,
$$\left|\frac{a-x_1}{a-x_2}\right| = \rho^2.$$

Hence a,—which was determined uniquely by the construction

when x was given,—is fixed for all positions of x for which ρ is constant. Also $|x-a|$ is constant, since

$$|x-a|^2 = |x_1 - a|\,|x_2 - a|;$$

therefore x is any point of a circle whose centre is a and radius $|\sqrt{(a-x_1)(a-x_2)}|$.

Thus the locus of x when it moves subject to the condition

$$\left|\frac{x-x_1}{x-x_2}\right| = \text{a constant},$$

is a complete circle.

The converse theorem is true. For if we take two points x_1, x_2, which lie on a ray from the centre a of a circle of radius ρ and which satisfy the equation

$$|a-x_1|\,|a-x_2| = \rho^2,$$

then if x be any point on this circle, $\left|\dfrac{x-x_1}{x-x_2}\right| = \text{a constant}.$

When the constant is 1 the circle of centre a degenerates into a straight line and the points x_1, x_2 are reflexions of each other in this line.

In the general case the points x_1, x_2 are said to be *inverse* as to the circle of centre a and radius ρ; and *the circle will be said to be drawn about any such pair of points.*

Any circle through two points is orthogonal to (i.e. cuts at right angles) any circle about those points. For let x_1, x_2 be the two points and let us draw (as in fig. 16) the circle x_1x_2x, then the join of a and x is a tangent to this circle at x by virtue of the equation

$$|x-a|^2 = |x_1 - a|\,|x_2 - a|.$$

It follows that the intersections of two circles which are orthogonal to a given circle are inverse points of that circle. This provides us with another definition for inverse points:—

Two points x_1, x_2 are said to be inverse to a circle, when every circle through x_1, x_2 cuts the given circle orthogonally.

22. The Bilinear Transformation converts Circles into Circles. Returning to equation (2), namely

$$\frac{y-y_1}{y-y_2} = k\frac{x-x_1}{x-x_2},$$

we see that $$\left|\frac{y-y_1}{y-y_2}\right| = |k| \left|\frac{x-x_1}{x-x_2}\right|.$$

This equation shows that when x describes a circle *about* x_1, x_2, or, what comes to the same thing, when x moves so that

$$\left|\frac{x-x_1}{x-x_2}\right| = \text{a constant } \rho,$$

then y moves so that

$$\left|\frac{y-y_1}{y-y_2}\right| = |k|\,\rho = \text{a constant};$$

that is, y describes a circle *about* y_1, y_2.

But x_1, x_2 are arbitrary; therefore *circles in the x-plane map into circles in the y-plane, and inverse points as to a circle map into inverse points as to the corresponding circle.*

But if $\rho = 1$ the x-path is a straight line and not a circle, and if $|k|\,\rho$ is 1 the y-path is a straight line.

We cannot say, with Euclid's definition of a circle, that a straight line is a circle, but we can say that it is the limit of a circle. It is however usual to say unreservedly that a circle maps into a circle when the transformation is bilinear; the limitation that either circle may be a straight line being implied.

Ex. 1. Prove that the map of a circle through x_1, x_2 is a circle through y_1, y_2, by equating amplitudes in equation (2).

Ex. 2. Let $y = \dfrac{2x+3}{3x-2}$; and let x take the values 0, $\pm i$, $\pm 2i$. Mark the corresponding points y; and determine the centre and radius of the y-circle.

23. Coaxial Circles. All circles *through* two points x_1, x_2 are said to be *coaxial*; and also all circles *about* the two points are said to be coaxial. The two systems have been called *hyperbolic* and *elliptic* coaxial systems; just as in projective geometry a conic is a hyperbola when it meets the line infinity in distinct real points, and an ellipse when it does not meet that line in real points. So also a system of circles touching at a point is a *parabolic* coaxial system. It is to be noticed that in the hyperbolic system we are concerned primarily with arcs whose ends are the points x_1, x_2, not with the complete circles.

The figure exhibits an elliptic and a hyperbolic system of coaxial circles.

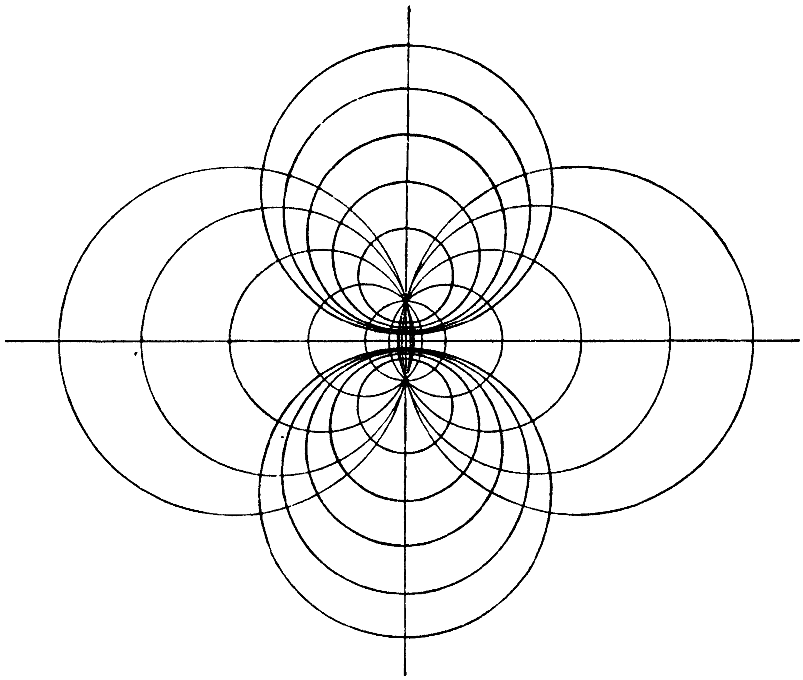


Fig. 17.

It is evident from the remarks of the last paragraph that *any coaxial system of circles in the x-plane, whether elliptic, hyperbolic or parabolic, maps into another such system in the y-plane, when the transformation is bilinear.*

Ex. Draw two orthogonal parabolic systems of coaxial circles.

24. Harmonic Pairs of Points. Let E be a circle about x_1, x_2, and H a circle through x_1, x_2; let E, H intersect at x_1', x_2'. We shall consider the relation of x_1', x_2' to x_1, x_2.

In studying the relation of a configuration and its map we look out for those properties of the one which reappear unaltered in the other. These properties are said to be *invariant.* Thus four points on a circle, say a cyclic tetrad, have an invariant property, for they map into four points on a circle. And, again, two points on a circle and two points inverse

as to that circle, say anticyclic pairs, have an invariant property. In our case the four points $x_1x_2x_1'x_2'$ are on a circle H, and two

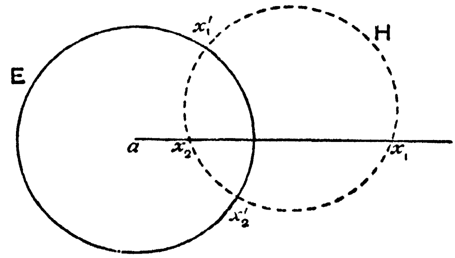


Fig. 18.

are on E, while two are inverse as to E; thus the four points,—or rather the two pairs,—are both cyclic and anticyclic.

In fig. 18 the two pairs (x_1, x_2), (x_1', x_2') lie on the circle H, and interlace so that the points of one pair are separated by the points of the other pair. Hence

$$\operatorname{am}\frac{x_1'-x_1}{x_2-x_1'} - \operatorname{am}\frac{x_2'-x_1}{x_2-x_2'} \equiv \pi.$$

Further, because x_1' and x_2' are on the circle E,

$$\left|\frac{x_1'-x_1}{x_2-x_1'}\right| = \left|\frac{x_2'-x_1}{x_2-x_2'}\right|.$$

Hence
$$\frac{x_1'-x_1}{x_2-x_1'} = -\frac{x_2'-x_1}{x_2-x_2'};$$

that is, the points x_1' and x_2' will divide the stroke from x_1 to x_2 in opposite ratios,—that is, in ratios of the same absolute value but of contrary signs.

Cleared of fractions the relation takes the form

$$(x_1+x_2)(x_1'+x_2') = 2(x_1x_2+x_1'x_2');$$

that is, the product of the sums is twice the sum of the products. In this form the symmetry of the arrangement is brought out. Not merely the points x_1 and x_2, or x_1' and x_2', enter symmetrically, but also the pairs x_1 and x_2, x_1' and x_2' enter symmetrically. The two pairs are said to be *harmonic*.

If we take the origin at the centroid of x_1, x_2, then $x_1+x_2=0$ and the relation takes the simple form

$$x_1'x_2' = -x_1x_3 = x_1^2 = x_2^2.$$

This relation gives

$$\frac{x_2'}{x_1} = \frac{x_1}{x_1'}, \quad \frac{x_2'}{x_2} = \frac{x_2}{x_1'},$$

and shows therefore that the strokes from the centroid of x_1, x_2 to the points x_1' and x_2' are inclined equally but oppositely to the stroke from x_1 to x_2.

Since the tangents at x_1' and x_2' to H meet on the join of x_1 and x_2, the joins of the pairs are *conjugate* lines as to the circle; this fact is fundamental in the transition from the present point of view to that of projective geometry[1].

Ex. 1. Prove that when x_1, x_2 and x_1', x_2' are harmonic,

$$\frac{2}{x_1 - x_2} = \frac{1}{x_1 - x_1'} + \frac{1}{x_1 - x_2'}.$$

Ex. 2. Taking three circles any two of which are orthogonal, prove that the two pairs of intersections which lie on any circle are harmonic.

NOTE. Thus we have three pairs of points any two of which are harmonic.

Ex. 3. In a regular hexad point out the harmonic pairs.

25. The Double Ratios of Four Points. Returning to the equation (2) (§ 21), namely,

$$\frac{y - y_1}{y - y_2} = k\frac{x - x_1}{x - x_2},$$

let x_3, y_3 and x_4, y_4 be corresponding points. Then

$$\frac{y_3 - y_1}{y_3 - y_2} = k\frac{x_3 - x_1}{x_3 - x_2},$$

$$\frac{y_4 - y_1}{y_4 - y_2} = k\frac{x_4 - x_1}{x_4 - x_2}.$$

Hence by division

$$\frac{y_3 - y_1}{y_3 - y_2} \cdot \frac{y_4 - y_2}{y_4 - y_1} = \frac{x_3 - x_1}{x_3 - x_2} \cdot \frac{x_4 - x_2}{x_4 - x_1} \quad\text{..............}(1).$$

We have then on the right side of this equation a quantity depending on four points which is unchanged by the bilinear

[1] The transition is effected by means of the circular points at infinity. In the present case the condition that two pairs of points are harmonic is the condition that the pairs of straight lines from them to either circular point are harmonic.

transformation. The quantity is called the *double ratio*, or anharmonic ratio, of the pairs x_3, x_4 and x_1, x_2, since $\frac{x_3-x_1}{x_3-x_2}$ is, with sign changed, the ratio in which x_3 divides x_1, x_2 (§ 18). If we denote the double ratio by $(x_3, x_4; x_1, x_2)$ we have

$$(x_4, x_3; x_1, x_2) = 1/(x_3, x_4; x_1, x_2),$$
$$(x_3, x_4; x_2, x_1) = 1/(x_3, x_4; x_1, x_2),$$
$$(x_4, x_3; x_2, x_1) = (x_3, x_4; x_1, x_2) = (x_1, x_2; x_3, x_4).$$

The double ratio depends, then, on the way in which we pair off the four points, and the way in which we associate one of one pair with one of the other. In the above double ratio we have first selected the pairs x_1, x_2 and x_3, x_4; and then associated x_3 with x_1, x_4 with x_2 to get $(x_3, x_4; x_1, x_2)$. We see then that since there are three ways of pairing off, and two ways of associating selected pairs, there are six double ratios of four points.

If we write

$$(x_2-x_3)(x_1-x_4)=l,\ (x_3-x_1)(x_2-x_4)=m,\ (x_1-x_2)(x_3-x_4)=n,$$

so that $$l+m+n=0,$$

then the six are

$$-m/n,\ -n/l,\ -l/m;$$
$$-n/m,\ -l/n,\ -m/l.$$

Special relations of four points, which are not alterable by a bilinear transformation, will betray themselves by special values of the double ratios. It is convenient in detecting these to suppose a transformation effected by which one of the points passes to ∞. Suppose y_4 is ∞, then the double ratios are

$$\frac{y_1-y_2}{y_1-y_3},\ \frac{y_2-y_3}{y_2-y_1},\ \frac{y_3-y_1}{y_3-y_2},$$

and their reciprocals.

Any one of these determines the shape of the triangle $y_1y_2y_3$. When one is real the triangle is flattened out, and all are real. Hence in general:

When one double ratio is real, all are real and the four points lie on a circle.

Or again, if a double ratio has the absolute value 1, then the triangle is isosceles; that is, two points are on a circle about the third and ∞. Therefore, when a double ratio is a mere turn, the pairs corresponding to that ratio are anticyclic.

There are two especially simple arrangements. (1) When the triangle is both flat and isosceles; the four points then pair off into harmonic pairs (§ 24) and the six double ratios reduce to three, namely, $-1,\ 2,\ 1/2.$
(2) When the triangle is *regular* or equilateral; the points then pair off in three ways into anticyclic pairs; and the six double ratios reduce to two, for the ratio $\dfrac{y_1-y_2}{y_1-y_3}$ is either $-v$ or $-v^2$.

Observe that in equation (1) by regarding x_4 and y_4 as variable, we have an equation which maps the x-plane on the y-plane when three corresponding pairs of points are assigned.

Ex. 1. Find the six double ratios of the points $0, 1, \infty, x$.

Ex. 2. If $(x, x_1; x_2, x_3) = -v$, find x.

Ex. 3. From the equation $l+m+n=0$ deduce that the sum of the rectangles of opposite sides of a convex quadrilateral is never less than the rectangle of the diagonals; and that it is equal to this rectangle when the points are cyclic.

Ex. 4. Determine the equation between x and y which maps $0, 1, \infty$ into $1, v, v^2$.

26. Isogonality. The derivate of y as to x is defined (just as for real variables) as the limit, if there be one, of the ratio of the change of y to the corresponding change of x, when the change of x is made arbitrarily small.

In the case of the bilinear transformation, we have

$$y_1 - y = \frac{(ad-bc)(x_1-x)}{(cx+d)(cx_1+d)};$$

therefore
$$\lim_{x=x_1} \frac{y_1-y}{x_1-x} = \frac{ad-bc}{(cx+d)^2}.$$

In fact the well-known rules for obtaining derivates apply unaltered to this case, so that we can write at once

$$D_x y = \frac{ad-bc}{(cx+d)^2}.$$

The derivate depends on x alone, not on the change of x. Let these two points x_1, x_2 approach x by different paths,—to fix ideas let it be along two circles,—and let y_1, y_2 be the corresponding points that are given by the bilinear transformation. Then

$$\lim \frac{y_1 - y}{x_1 - x} = \lim \frac{y_2 - y}{x_2 - x} = D_x y,$$

so that, if $D_x y$ be neither 0 nor ∞,

$$\lim \frac{y_1 - y}{y_2 - y} = \lim \frac{x_1 - x}{x_2 - x}.$$

But the amplitude on the left is the angle which the path of y_1 makes with the path of y_2. Therefore *the angle which one arc through y makes with any other is equal to the corresponding angle in the x-plane ; or again, when x, instead of keeping to one circle, turns and moves on another circle, then y turns through the same angle* (fig. 19).

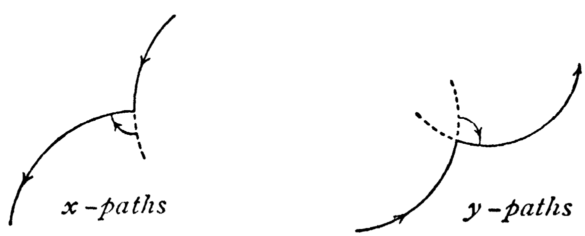


Fig. 19.

What has been said of circles may be said of other corresponding paths which are such that the tangents at the points of intersection are determinable.

We have excepted the cases when $D_x y$ is 0 or ∞; that is, when x or y is ∞. And the theorem has then on the face of it no meaning, as we have not assigned any meaning to an angle at ∞. But since two straight lines through ∞ are to be regarded as two special circular arcs, and one circular arc makes opposite angles with another at their two intersections, we regard the angle at ∞ which a stroke from x to ∞ makes with another from x to ∞ as the opposite of the angle made at x. With this understanding the property of isogonality, as it is called, is true always for the transformation $y = (ax + b)/(cx + d)$.

The property of isogonality applies to other correspondences than the 1, 1 correspondence; only in these higher correspondences there are exceptional points at which a modification is necessary.

The study of the behaviour of x, y at these exceptional points in the y, x planes forms a very important part of the theory of functions..

27. Theory of Absolute Inversion. Let us consider now the special bilinear transformation for which $a = d = 0$, $b = c = 1$; that is, the transformation $y = 1/x$. Let

$$x = \rho(\cos\theta + i\sin\theta),\ y = \rho'(\cos\theta' + i\sin\theta').$$

Then
$$\rho' = 1/\rho \text{ and } \theta' = -\theta.$$

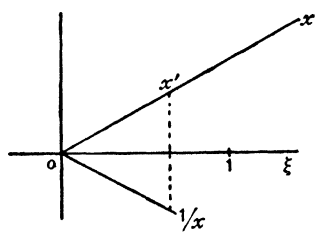


Fig. 20.

Thus the point $1/x$ can be constructed by (1) taking the inverse x' of x as to the unit circle with centre o, and (2) reflecting x' in the ξ-axis; (2) can be regarded as taking the inverse of x' as to the ξ-axis (§ 21). Thus we can obtain $1/x$ from x by two inversions; which can moreover be performed in any order. We shall show that in general the bilinear transformation is equivalent to two inversions; but we shall first discuss briefly the principles of this method of inversion. Inasmuch as the substitution of a/x for x/a may itself be called an inversion, we may call, to avoid confusion, the substitution of $|a/x|$ for $|x/a|$ an *absolute inversion.*

Take a fixed sphere whose centre is O and radius α; and let P, Q be two points on a line from O such that

$$OP \,.\, OQ = \alpha^2;$$

then P, Q are *inverse points* of the sphere. By this means the

points inside the sphere are paired off one-to-one with the points outside. Points on the sphere are self-correspondent. Call this sphere the sphere of inversion.

I. We shall examine first the effect of inversion upon the

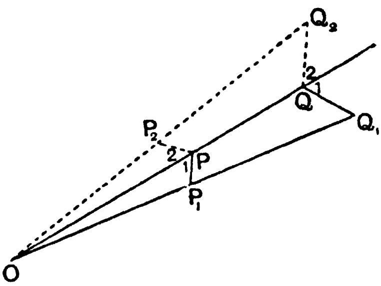


Fig. 21.

angle of intersection of two curves. Considering the two pairs PQ and P_1Q_1, we have

$$OP \,.\, OQ = OP_1 \,.\, OQ_1,$$

an equation which shows that P, P_1, Q_1, Q lie on a circle. Hence the angles OPP_1 and OQ_1Q are equal in magnitude. Taking P_1 near P, and therefore Q_1 near Q, we see that when P moves in a direction making an angle θ with OP, Q moves in the same plane and its path makes with OQ the angle $\pi - \theta$; for the angles marked 1 are ultimately equal.

So also taking a third pair P_2Q_2, the angles 2 are ultimately equal. Hence P_1PP_2, Q_1QQ_2 are ultimately equal.

Taking P_1 and P_2 near P, and therefore Q_1 and Q_2 near Q, we can regard PP_1, PP_2 and QQ_1, QQ_2 as elements of tangents to two curves at P and to two inverse curves at Q; we have proved then that *the two curves inverse to two assigned curves intersect at an angle whose magnitude is that of the angle of intersection of the original curves.*

Suppose that the curves lie in a common plane, so that OPQ, OP_1Q_1, OP_2Q_2 lie in a plane. The angles of intersection are of the same size but opposite directions (fig. 22). Thus in this case *when a curve C' makes with a curve C an angle θ at a point P, the inverse of C' makes with the inverse of C the angle $-\theta$ at the point Q, inverse to P.*

Briefly we may say that the inverse of an angle is the opposite angle.

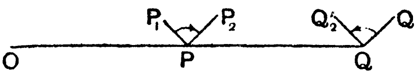


Fig. 22.

II. We proceed next to the proof of the fundamental theorem that the inverse of a sphere is a sphere. To make the statement universal we regard a plane as a sphere passing through ∞. Observe that the 1, 1 correspondence of points is to hold always; but when P is at O, Q is at ∞, hence we are led to regard ∞ as a point in this kind of geometry.

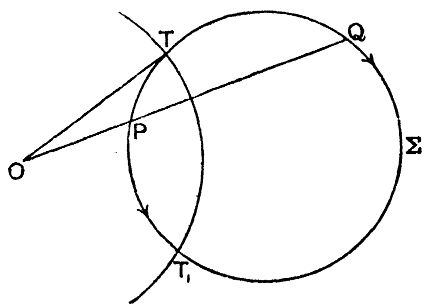


Fig. 23.

(i) When the radius of the sphere of inversion is the length OT (fig. 23), the sphere Σ inverts into itself; since

$$OP \, . \, OQ = OT^2.$$

Thus *any sphere orthogonal to the sphere of inversion inverts into itself.*

(ii) Next let OP meet any sphere Σ at Q and R.

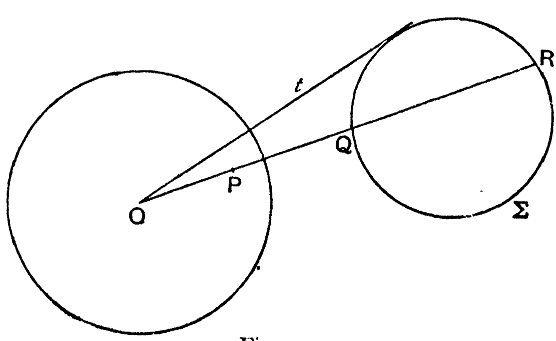


Fig. 24.

Then $$OP \, . \, OQ = a^2,$$
$$OQ \, . \, OR = t^2,$$
where t is the length of the tangent from O to Σ.

Therefore $$OP \propto OR,$$
and P lies on a sphere. Hence *the inverse of a sphere that passes through neither O nor ∞ is a sphere.*

(iii) But if Σ passes through O, let OA be a diameter of Σ and let OP intersect the tangent plane at A in R.

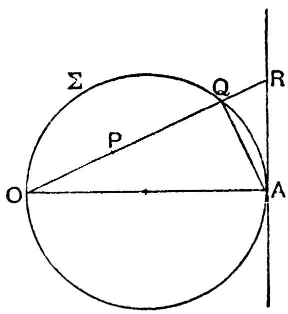


Fig. 25.

Then $$OP \, . \, OQ = a^2,$$
$$OQ \, . \, OR = OA^2;$$
therefore $$OP \propto OR,$$
and P lies on a plane parallel to the tangent plane at A. Hence *the inverse of a sphere through O is a sphere through ∞.*

(iv) Lastly if the sphere passes through O and ∞, that is if it is a plane through O, the inverse is clearly the same plane. If we restrict our view to this plane we have the ordinary *plane inversion.*

III. We have proved that the inverse of a sphere is a sphere. It is now easy to prove that the inverse of a circle is a circle.

The points common to two surfaces invert into the points common to the inverse surfaces. Let the surfaces in question be two spheres that pass through an assigned circle; then the inverse spheres pass through a circle which is inverse to the

given circle. In particular we see that *in plane inversion a circle inverts into a circle.*

Inversion with respect to a plane. When the sphere of inversion becomes a plane, what is to be understood by *inverse points?*

In the general case let OP meet the sphere at X, Y. Then

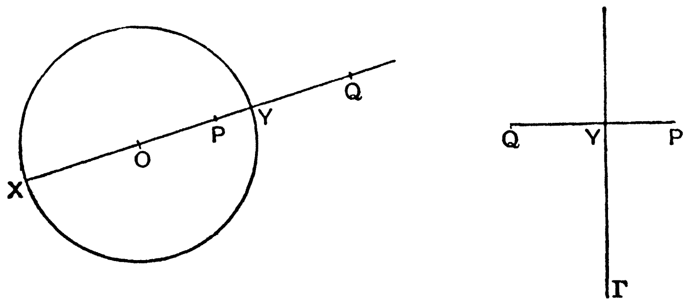


Fig. 26.

X, Y are harmonic with P, Q. Thus when O and therefore X passes to infinity, Y being fixed,

$$PY = YQ.$$

Hence P and Q are reflexions in the plane (fig. 26).

28. The Bilinear Transformation is equivalent to Two Inversions in Space. The relation $y = (ax + b)/(cx + d)$ involves four constants, but these constants appear only as ratios; hence a knowledge of three pairs of corresponding values (x, y) will determine completely the character of the transformation. As corresponding values let us take

$x =$	∞	$-d/c$	$-b/a$
$y =$	a/c	∞	$0.$

Taking any point X exterior to the x-plane as origin, invert the x-plane into a sphere through X. The correspondence between the points of the sphere and the values of x is one-to-one; in particular the value $x = \infty$ corresponds to the point X itself. Let the points $-d/c$, $-b/a$ of the plane pass into points Y, Z of the sphere; and invert the sphere again, this time from Y, on a plane parallel to the tangent plane at Y and at such a

distance that Z, X invert into points whose distance apart is $\left|\frac{a}{c}\right|$. Taking this plane as the y-plane and the inverse of Z as the origin, we can adjust the real axis in this plane so that the inverse of X is a/c.

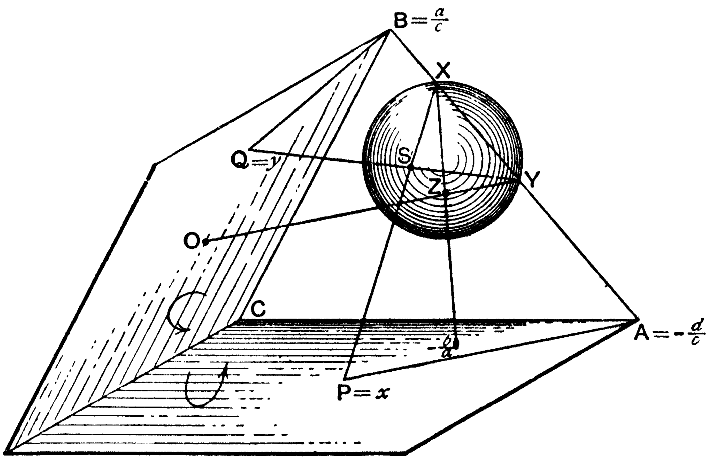


Fig. 27.

Now let any point S of the sphere invert into the points P, Q on the two planes, the centres of inversion being X and Y respectively; and let P be the point x of the x-plane, Q the point y of the y-plane.

Then
$$\frac{AP}{XA} = \frac{SY}{SX} = \frac{BY}{BQ},$$

or
$$AP \,.\, BQ = XA \,.\, BY = \text{a constant},$$

and taking for P the particular position $-b/a$, the constant is

$$\left|-\frac{b}{a}+\frac{d}{c}\right| \left|0-\frac{a}{c}\right|, \text{ or } \left|\frac{bc-ad}{c^2}\right|.$$

Hence
$$\left|x+\frac{d}{c}\right| \left|y-\frac{a}{c}\right| = \left|\frac{bc-ad}{c^2}\right|.$$

We have proved then that the two quantities

$$\left(x+\frac{d}{c}\right)\left(y-\frac{a}{c}\right), \quad \frac{bc-ad}{c^2},$$

are equal in absolute value.

The plane through X, Y and the centre of the sphere will meet the line of intersection of the planes perpendicularly, say at C. Since CA, CB are respectively perpendicular to the diameters through X and Y, $CB = CA$. And since AP and BQ lie in a plane, the angles CAP, CBQ are equal in magnitude. Let us regard angles in the two planes as positive when they are in the senses shown by the arrows,—a natural convention since to an observer standing on the sphere at X angles in the x-plane now increase counter-clockwise, and when he moves to Y they still increase counter-clockwise. Then we have

$$\angle CAP + \angle CBQ = 0,$$

that is, $$\text{am}\left(x + \frac{d}{c}\right) + \text{am}\left(y - \frac{a}{c}\right) = \text{constant};$$

and taking $$x = -\frac{b}{a},\ y = 0,$$

we get $$\text{am}\left(x + \frac{d}{c}\right)\left(y - \frac{a}{c}\right) = \text{am}\left(-\frac{b}{a} + \frac{d}{c}\right)\left(-\frac{a}{c}\right)$$

$$= \text{am}\,\frac{bc - ad}{c^2}.$$

Thus $\left(x + \frac{d}{c}\right)\left(y - \frac{a}{c}\right)$ and $\frac{bc - ad}{c^2}$, which were proved equal in absolute value, are equal also in amplitude. That is, they are equal; hence,

$$y = \frac{ax + b}{cx + d}.$$

It is proved, then, that *the bilinear transformation is equivalent to two absolute inversions.*

Since we know that a circle inverts into a circle we are led again to the theorem that a circle maps into a circle when subjected to a bilinear transformation.

EXAMPLES.

1. Given three points a_1, a_2, a_3, construct the three b_1, b_2, b_3 such that a_1, b_1 and a_2, a_3 are harmonic, a_2, b_2 and a_3, a_1 are harmonic, and a_3, b_3 and a_1, a_2 are harmonic. Prove that the straight lines joining a_1 to b_1, a_2 to b_2, a_3 to b_3 meet at a point, and that the relation of the two triads is mutual.

2. Three points determine three circles, each circle passing through one point and about the other two. Prove that these circles meet at angles $\pi/3$ at two points.

3. Prove that two points on a circle subtend at any pair of inverse points angles whose sum is constant.

4. A point x is reflected in two plane mirrors. Prove that the second reflexion is the point x cis $2a$ where a is the angle which the second mirror makes with the first; the intersection of the mirrors being the origin.

5. Prove that a circle described positively in the x-plane maps into a circle described positively in the y-plane when it contains the point $-d/c$.

6. Taking four points a, b, c, d, prove that when any three are inverted as to a circle whose centre is the remaining point, the triangles formed by the inverse points are all similar.

CHAPTER IV

GEOMETRIC THEORY OF THE LOGARITHM AND EXPONENTIAL.

29. **Sketch of the Theory for Positive Numbers.** The theory of the logarithm is so important for the general theory of functions that it seems to us desirable to present it from two points of view; and in this chapter we shall give the geometric view, reserving for a later chapter the numerical view which is indispensable for calculation.

The way in which the exponential and logarithm are introduced in Algebra cannot well be utilized here, for the processes used are precisely those which we have to inquire into later. We make acquaintance with these processes in Algebra and subsequently we should inquire into the logic of them. But in the Calculus we rely on what was said in Algebra; and as the geometric function-theory requires such facts as $D \log \xi = 1/\xi$, where ξ is real, there is a dilemma. In the case of $\sin \xi$ there is no such difficulty; for we have first a geometric definition, and then a geometric proof that $D \sin \xi = \cos \xi$. For these reasons to put the logarithm on a similar basis, whereby we can make use of it without first discussing the theory of infinite series, or even the irrational exponent, is very desirable for elementary purposes.

We shall assume the existence of the equiangular spiral, which may be constructed by placing a right cone with axis vertical and attaching a thread AP to a point A of it. The point P is held in the horizontal plane through the vertex O;

the thread is then wound round the cone, without being allowed to slip. As the thread is not to lie in a geodesic on the cone, the latter should not be polished. The curve described by P in its plane is the spiral in question, and it can be proved in an elementary way to have this characteristic property: that it cuts all rays from O at a constant angle. Choose this angle to be $\pi/4$; and then measuring the angle θ from the point where $OP = \rho$ has the value 1, define θ as the *logarithm* of ρ:—

$$\theta = \operatorname{Log} \rho \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1).$$

Thence by a geometric limit-process, namely that which gives the formula $\rho D_\rho \theta = \tan \phi$, where ϕ is an angle made by the tangent with the radius vector, we have

$$\rho D_\rho \theta = 1,$$

and
$$D \operatorname{Log} \rho = 1/\rho.$$

We observe that for a given ρ there is but one θ; so that θ and $\theta + 2\pi$ are here quite distinct, defining not a direction but the amount of turning.

We define inversely ρ as the *exponential* of θ,

$$\rho = \exp \theta,$$

whence
$$\exp 0 = 1, \text{ and } D \exp \theta = \exp \theta.$$

Take two points of the spiral, ρ, θ and ρ_1, θ_1, and let $\theta_1 - \theta$ be a constant.

We have then $D_{\theta_1} \rho_1 = D_\theta \rho_1$,

$$D(\rho_1/\rho) = \frac{\rho D \rho_1 - \rho_1 D \rho}{\rho^2} = 0;$$

that is, when we keep the angle $\theta_1 - \theta$ constant, the ratio ρ_1/ρ is also constant.

From this we infer that ρ_1/ρ depends only on $\theta_1 - \theta$, and since

when
$$\rho = 1, \ \theta = 0, \text{ and } \rho_1 = \exp \theta_1,$$

therefore
$$\rho_1/\rho = \exp(\theta_1 - \theta),$$

and
$$\operatorname{Log}(\rho_1/\rho) = \theta_1 - \theta = \operatorname{Log} \rho_1 - \operatorname{Log} \rho.$$

We are now at the same point, with regard to the logarithm, as with regard to the sine before the series for $\sin x$ is proved; that is before we study analytic trigonometry.

There is one other parallelism to be mentioned. In trigonometry we speak of a constant π, without at first explaining any satisfactory way of calculating it. So here we can suppose known that value of ρ which corresponds to $\theta = 1$, that is $\exp 1$. A rough measurement shows that it is nearly 2·7; we denote it by e, and regard its calculation as belonging to analysis, where it is shown that

$$e = 2{\cdot}718281828\ldots.$$

It is geometrically evident that as ρ passes through all *positive* values θ passes through all *real* values.

Equation (1) shows that $\int d\rho/\rho$, taken from one value of the positive number ρ to another, is the change of $\operatorname{Log} \rho$.

30. The Logarithm in general. We proceed to the logarithm of any number.

Let x be $\rho \operatorname{cis} \theta$. Then

$$dx = d\rho \operatorname{cis} \theta + \rho d \operatorname{cis} \theta.$$

But $\quad d \operatorname{cis} \theta = (-\sin \theta + i \cos \theta)\, d\theta = i \operatorname{cis} \theta \, d\theta;$

hence $\quad dx = (d\rho + i\rho d\theta) \operatorname{cis} \theta,$

and $\quad dx/x = d\rho/\rho + i d\theta.$

There is then this peculiarity about dx/x as compared with $x^n dx$ in general,—that it separates into a real part involving only ρ, and an imaginary part involving only θ. Thus before discussing integrals of expressions involving x in general we can discuss this very important case; for we have, if initially $x = 1$, $\rho = 1$, $\theta = 0$,

$$\begin{aligned} \int_1^{x_1} dx/x &= \int_1^{\rho_1} d\rho/\rho + i \int_0^{\theta_1} d\theta \\ &= \operatorname{Log} \rho_1 + i\theta_1 \\ &= \operatorname{Log} |x_1| + i \operatorname{am} x_1. \end{aligned}$$

This expression, $\operatorname{Log} \rho_1 + i\theta_1$, we call the logarithm of x_1, or $\log x_1$.

The first term is the natural logarithm of ρ_1,—a real number, defined in the last article. It depends only on the distance ρ_1 of the final x from the origin. The second term $i\theta_1$ depends on

the path by which x has passed from 1 to x_1—that is on *the path of integration.* It depends on the amount of turn of the stroke x in passing from 1 to x_1. Thus $\log x$ is many-valued in just the same way as the amplitude. According to the path selected it is capable of an arithmetic series of values differing by $2\pi i$; and the general value of $\log x_1$ is found by adding $2n\pi i$ to any particular value.

Ex. Find the values of $\log 1$, $\log i$, $\log -i$, $\log v$.

In the integration of dx/x we have had to pay attention to the route from the initial to the terminal point. When the variable is real the path of integration of a definite integral is determinate and unique, for the variable is confined to the real axis and must increase or decrease constantly from the initial to the final value. But when the variable can take complex values there are as many (i.e. infinitely many) ways of passing continuously from the one limit of the integral to the other as there are ways of connecting the corresponding points in the plane.

We define that value of $\log x$, for which $-\pi < \theta \leqq \pi$, as the *chief value*, or *chief branch*, of the logarithm; and denote it by $\operatorname{Log} x$. Thus $\operatorname{Log} x = \operatorname{Log} \rho + i \operatorname{Am} x$.

We have $\operatorname{Log}(\rho_1/\rho_0) = \operatorname{Log}\rho_1 - \operatorname{Log}\rho_0$;

so that

$$\begin{aligned}\log(x_1/x_0) &= \operatorname{Log}\rho_1/\rho_0 + i(\theta_1 - \theta_0)\\ &= \operatorname{Log}\rho_1 + i\theta_1 - (\operatorname{Log}\rho_0 + i\theta_0)\\ &= \log x_1 - \log x_0 + 2n\pi i.\end{aligned}$$

In the same way it is proved at once that the logarithm of a product is congruent (to the modulus $2\pi i$) with the sum of the logarithms of the factors.

Since (§ 15) $\operatorname{Am}(x_1/x_0)$ is not always $\operatorname{Am} x_1 - \operatorname{Am} x_0$, it is not always true that

$$\operatorname{Log}(x_1/x_0) = \operatorname{Log} x_1 - \operatorname{Log} x_0.$$

In the first chapter we discussed two essentially different ways of comparing two strokes or complex numbers x_0 and x_1. The one way is to consider the difference, or speaking physically the displacement; the other is to consider the ratio, or the stretch and turn. And the logarithm may

be regarded as affording the necessary means of transition from the one to the other; for when x_0 becomes x_1 by a turn $\theta_1 - \theta_0$ and a stretch ρ_1/ρ_0, then the logarithm, y suppose, makes in its plane a step $\operatorname{Log} \rho_1/\rho_0$ eastward and a step $\theta_1 - \theta_0$ northward. Observe how particularly simple is the case when there is no stretch. Then $\rho_1 = \rho_0$, and the logarithm is merely $i \times$ turn. Whereas in the case where there is no turn the logarithm, though real, can be assigned to a close degree of approximation only by a complicated calculation. With this calculation the reader is already familiar from algebra, where series, such as

$$2 \operatorname{Log} \rho = \frac{\rho - 1}{\rho + 1} + \frac{1}{3}\left(\frac{\rho - 1}{\rho + 1}\right)^3 + \dots,$$

are given. The validity of this series for all positive values of ρ will appear incidentally later on.

31. Mapping with the Logarithm. Let now $y = \log x$, or $\xi' + i\eta' = \operatorname{Log} \rho + i\theta$.

We have then $\xi' = \operatorname{Log} \rho, \quad \eta' = \theta.$

Thus to the circles $\rho = \text{const.}$ in the x-plane correspond the lines $\xi' = \text{const.}$ in the y-plane; to the circles of radii $1, e, e^2, \dots, e^{-1}, e^{-2}, \dots$, correspond equidistant lines $\xi' = 0, 1, 2, \dots, -1, -2, \dots$. To the rays $\theta = \text{constant}$ in the x-plane correspond the lines $\eta' = \theta$ in the y-plane; and if the rays are drawn at equal angles the lines are equidistant. Figure 28 shows the mapping.

As θ increases, the line $\eta' = \theta$ moves upwards; and when θ is 2π, the map is not $\eta' = 0$ again, but $\eta' = 2\pi$. Thus to one ray in the x-map correspond infinitely many equidistant straight lines in the y-map.

If we restrict θ by the condition

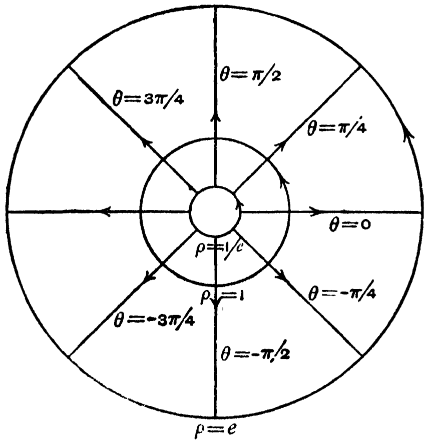
$$-\pi < \theta \leqq \pi,$$

that is if we consider only the chief logarithm $\operatorname{Log} x$, then we have the whole of the x-plane, but the corresponding region in the y-plane is bounded by the lines $\eta' = -\pi$ and $\eta' = \pi$, and includes the latter line but not the former.

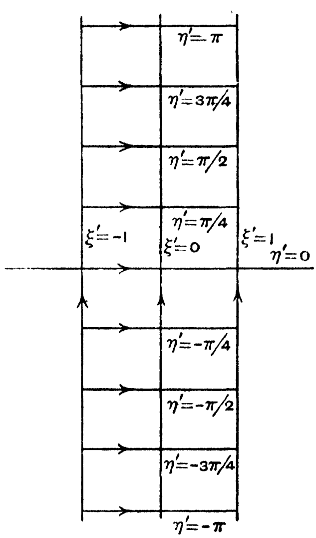
It will be noticed that orthogonal curves map into orthogonal

curves, in accord with the property of isogonality (§ 26). For we have here

$$D_x y = 1/x.$$



x - plane.



y - plane.

Fig. 28.

We have therefore isogonality except when $x = 0$ or when $x = \infty$

We can use this property to determine the curve in the x-plane which will map into a straight line in the y-plane. For the curve must cut all the rays at the same angle; it is therefore

an equiangular spiral. To obtain its polar equation, let the line in the y-plane be

$$\xi' = \alpha\eta' + \beta,$$

where α, β are constants.

Then $$\operatorname{Log} \rho = \alpha\theta + \beta,$$

which is the equation required.

The angle of the spiral is the angle which the y-line makes with the lines $\eta' = \text{const.}$, and is the same for all parallel y-lines.

The orthogonal systems in the x-plane are two specially placed coaxial systems of circles. It may be left to the reader to show that the chess-board arrangement in the y-plane is two specially placed parabolic coaxial systems of circles (§ 23).

32. The Exponential. When y is the logarithm of x, we call x *the exponential* of y; and write $x = \exp y$. The value of x is determined by the equations

$$\rho = \exp \xi', \quad \theta = \eta';$$

the exponential of the real number ξ' has been already defined in § 29. Thus for a given y there is but one x, so that the exponential of a complex number is one-valued. But we know that when x corresponds to y it also corresponds to $y + 2n\pi i$; that is $\exp y$ and $\exp(y + 2n\pi i)$ are the same whatever integer n may be. A function which repeats its value when the argument increases by a fixed amount is called *periodic*; thus $\exp y$ is periodic and has the *period* $2\pi i$; just as $\sin y$ is periodic, but with the real period 2π.

When y is imaginary,—that is when $\xi' = 0$,—we have $\rho = 1$, and x is $\operatorname{cis}\theta$, while y is $i\eta'$ or $i\theta$. Thus our definition of the exponential leads us to say that

$$\operatorname{cis}\theta \text{ is } \exp i\theta;$$

in other words that

$$\exp i\theta = \cos\theta + i\sin\theta.$$

When we can assert, from the analytic theory (ch. XII.), that

$$\exp x = 1 + x + x^2/2! + x^3/3! + \dots$$

for any assigned x, then we have

$$\cos\theta + i\sin\theta = 1 + i\theta + (i\theta)^2/2! + (i\theta)^3/3! + \dots,$$

and
$$\cos\theta = 1 - \theta^2/2! + \theta^4/4! - \dots,$$
$$\sin\theta = \theta - \theta^3/3! + \theta^5/5! - \dots,$$
the analytic definitions of the cosine and the sine.

In seeking to connect $\log x$ or $\exp x$ with x itself we must have in any case the problem of a *limit*. The limit employed in this chapter is $\int d\rho/\rho$ taken from one value of the positive quantity ρ to another.

We shall see, in the analytic theory, that
$$\lim_{n=\infty} n\,(x^{1/n} - 1) \text{ is } \operatorname{Log} x,$$
where $x^{1/n}$ is the chief nth root of x. It is well to verify this limit geometrically when $x = \operatorname{cis}\theta$. To find the limit in this case we observe that, θ_0 being the chief amplitude, $(\operatorname{cis}\theta)^{1/n} = \operatorname{cis}(\theta_0/n)$, and that $\operatorname{cis}(\theta_0/n) - 1$ is the stroke from 1 to a point of the unit

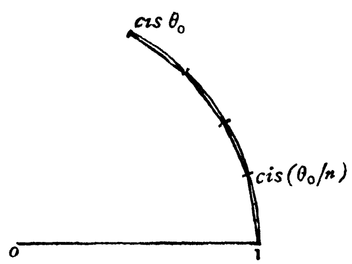


Fig. 29.

circle (fig. 29) whose arcual distance from 1 is $1/n$ of the arcual distance from 1 to x.

Thus $n\,(\operatorname{cis}(\theta_0/n) - 1)$, as to absolute value, is the sum of the n chords of n arcs whose sum is θ_0; while its amplitude is $\pi/2 + \theta_0/2n$. Hence the limit has the absolute value θ_0 and the amplitude $\pi/2$; that is, it is $i\theta_0$.

33. Mercator's Projection. The application of the preceding principles to the problem of making maps of the Earth's surface is of great interest. The discovery of the compass brought with it the idea of steering a course which should make with all meridians a constant angle. This spiral course was called a loxodrome or rhumb line. When the Earth's surface (regarded as a sphere) is inverted from the north pole into a

plane, say into the tangent plane at the south pole, the meridians become a system of rays in the plane, and the loxodromes become, by isogonality, curves cutting these rays at a constant angle, or equiangular spirals. Now the loxodromes being the important lines, the map so formed by 'stereographic projection' was not sufficiently simple; what was wanted was a map in which the loxodromes should appear as straight lines. This is done by mapping the inverse of the sphere by means of $y = \log x$; this is the principle of Mercator's projection.

34. The Addition Theorem of the Exponential. We know that if

$$x_3 = x_1 x_2,$$

then

$$\log x_3 = \log x_1 + \log x_2 + 2n\pi i.$$

Let $x = \exp y$, and let y_1, y_2, y_3 correspond to x_1, x_2, x_3; we see that

if

$$\exp y_3 = \exp y_1 . \exp y_2,$$

then

$$y_3 = y_1 + y_2 + 2n\pi i.$$

Therefore

$$\exp (y_1 + y_2 + 2n\pi i) = \exp y_1 . \exp y_2;$$

or

$$\exp (y_1 + y_2) = \exp y_1 . \exp y_2.$$

So

$$\exp (y_1 - y_2) = \exp y_1 / \exp y_2.$$

Ex. Write $\exp \pi i$, $\exp (\pi i/2)$, $\exp \frac{\pi}{1+i}$, as complex numbers of the form $\xi + i\eta$.

35. Napierian Motion. If we consider a moving point x as depending on the time t, then the derivate of x as to t, which we denote by $\dot{x}$, is the *velocity* of x at the time t; this velocity being itself a stroke whose absolute value, or magnitude, is the speed, and whose direction is along the tangent of the path of x. If x is $\rho \operatorname{cis} \theta$, then

$$\dot{x} = (\dot{\rho} + i\rho\dot{\theta}) \operatorname{cis} \theta,$$

so that the velocity is the sum of velocities of magnitude $\dot{\rho}$ in the direction θ, and magnitude $\rho\dot{\theta}$ in a perpendicular direction.

The simplest supposition we can make as to the velocity of a point is that it is constant; and the simplest supposition we can make as to the motion of all the points of a plane is that they all move with the same constant velocity,—this is a

translation. We suppose any point of the plane to have the initial position x_0 when the time t is 0, and to take the position

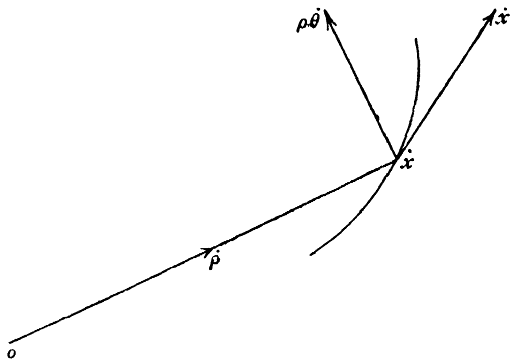


Fig. 30.

x_t at any time t; then on the above supposition we have $\dot{x}_t = b$, a given complex number.

Hence $$x_t = x_0 + bt.$$

If then we are considering a substitution $x_1 = x_0 + b$, we can suppose the new arrangement brought about by a translation; x_0 being the position of any moving point at time 0, x_1 at time 1. The motion of translation is of course only one of infinitely many ways in which the substitution can be brought about, but it is the simplest.

Suppose now that we impose on the y-plane a translation such that every point y_0 becomes $y_0 + bt$ at time t; this imposes on the x-plane a motion such that the corresponding x_0 (or $\exp y_0$) becomes x_t, where

$$x_t = \exp(y_0 + bt) = \exp y_0 \exp bt = x_0 \exp bt.$$

Thus at any given time the ratio x_t/x_0 is constant. The paths, or lines of flow, in the y-plane are parallel straight lines; hence the corresponding paths, or lines of flow, in the x-plane are equiangular spirals with a common pole $x = 0$ and with the same constant angle. Also those points of the y-plane which lie on a straight line at any instant will lie on a straight line at any other instant; taking in particular the lines of level, which are orthogonal to the lines of flow in the y-plane, these will map into another set of equiangular spirals in the x-plane orthogonal to the lines of flow; these also are lines of level. The lines of

flow and of level are shown in fig. 31, one system of lines being shown by dots. Either system being taken as lines of flow, the other will be lines of level.

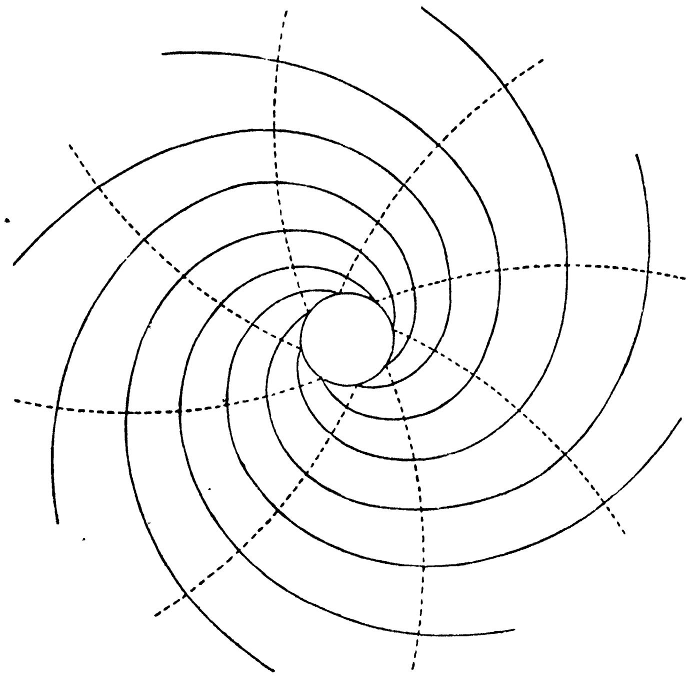


Fig. 31.

The velocity along a spiral is determined from the equations

$$\rho = \exp \xi', \quad \theta = \eta'.$$

These give $\dot{\rho}/\rho = \dot{\xi}'$, $\dot{\theta} = \dot{\eta}'$, and $\dot{\xi}'$, $\dot{\eta}'$ are respectively the constant eastward and northward velocities in the y-plane. Thus in the spiral motion the velocity along the radius ρ is proportional to that radius and the angular velocity $\dot{\theta}$ is constant.

In particular corresponding to a northward translation in the y-plane is a motion in the x-plane for which $\dot{\rho}$ is 0, $\dot{\theta}$ is constant; this is a rotation of the x-plane about the origin. And corresponding to an eastward translation in the y-plane is a motion in the x-plane for which $\dot{\theta}$ is 0, $\dot{\rho} \propto \rho$; this is a special dilatation of the plane, of the kind contemplated by Napier in introducing logarithms. In general the motion of the x-plane is a superposition of a rotation and a dilatation. For convenience this general motion will be called *Napierian*.

CHAPTER V.

THE BILINEAR TRANSFORMATION OF A PLANE INTO ITSELF.

36. The Fixed Points. The bilinear transformation

$$x_1 = (ax_0 + b)/(cx_0 + d)$$

converts a point x_0 of the x-plane into the point $(ax_0 + b)/(cx_0 + d)$; there is nothing to indicate that the transference is to take place along any special path. But, as in § 35, there is an advantage in connecting the substitution artificially with a special motion of the plane as a whole.

Let us begin with the simple form

$$x_1 = ax_0 + b.$$

There is one finite point which is unaltered, namely the point for which $x_1 = x_0$, or $ax_0 + b = x_0$; that is, the point $x_0 = b/(1 - a)$. Taking this point as origin, the equation becomes

$$x_1 = ax_0;$$

that is, all strokes are to be altered in the same ratio. This is attained by the Napierian motion.

If however $a = 1$ the above reduction fails; but the equation is then $x_1 = x_0 + b$ and only a translation is necessary (§ 35).

We consider in general the mapping of a plane on itself, by means of the equation

$$x_1 = (ax_0 + b)/(cx_0 + d) \qquad (1);$$

and regarding x_0 as the position of any point at the time $t = 0$, x_1 as the position of that same point at the time $t = 1$, we seek an appropriate motion of the plane.

There are two points which do not move, namely the points given by $x_1 = x_0$, or the points given by the equation

$$cx^2 + (d-a)x - b = 0.$$

Let this equation have, in the first place, two distinct roots f, f'; these roots are called the *fixed* (or *double*) points of the transformation.

With the help of these fixed points we can write (1) in the form

$$\frac{x_1 - f}{x_1 - f'} = k\frac{x_0 - f}{x_0 - f'} \quad\text{.....................}\quad (2),$$

for this equation is bilinear and shows, from its structure, that when x_0 is f (or f'), x_1 is f (or f'). The value of k can be found by letting x_1 become infinite; the value of x_0 is then $-d/c$, and we have

$$1 = k\frac{-d/c - f}{-d/c - f'},$$

or $$k = (cf' + d)/(cf + d). \qquad \text{(Compare § 21.)}$$

37. The Motion when the Fixed Points are distinct. If we write $z = (x-f)/(x-f')$, then (2) becomes

$$z_1 = kz_0.$$

Evidently the fixed points in the z-plane must correspond to those in the x-plane since $z = z_0$ when $x = x_0$. The new fixed points are 0 and ∞.

In this change of planes we have mapped on the z-plane not merely the old and new arrangements of the points of the x-planes but also the intermediate arrangements through which the totality of x-points may be supposed to reach their new positions x'. Thus we *consider the motion of the one plane as mapped into the motion of the other plane.*

Now when $z_1 = kz_0$, a Napierian motion suffices for the z-plane; from this we infer a suitable motion in the x-plane, by which x_0 becomes $(ax_0 + b)/(cx_0 + d)$. Let us examine this more closely.

(i) When $|k| = 1$ the motion of z is along a circle whose centre is O; hence the motion of x is along a circle *about* f and

f', and the lines of flow are an elliptic system of coaxial circles (§ 23). The lines of level are the orthogonal arcs of circles; that is all points which lie at any instant on one of these arcs lie at any other instant on another. *The substitution* (1) *is called elliptic in this case.*

(ii) When $\mathrm{Am}\, k = 0$, z moves along a ray from 0 to ∞; hence x moves along a circular arc between f and f'; the lines of flow are the system of such arcs; and the lines of level are the orthogonal system of circles. *This form of the substitution* (1) *is called hyperbolic.*

(iii) In general when $|k| \neq 1$ and $\mathrm{Am}\, k \neq 0$, z moves along an equiangular spiral; hence x moves along the map of such a spiral. What this map is like is easily seen from the property of isogonality, for it must cut all arcs between f and f' at the constant angle $\mathrm{Am}\, k$.

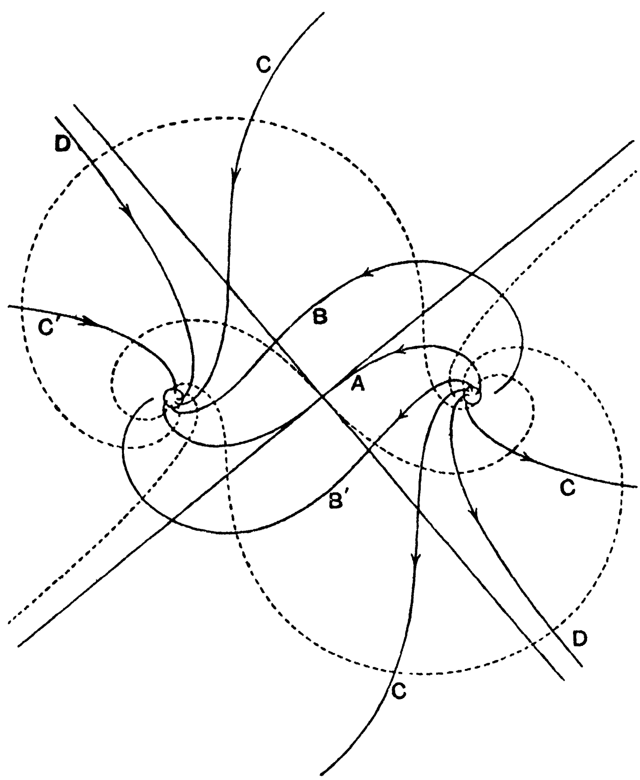


Fig. 32.

The curve is a double spiral, winding about the fixed points. The lines of flow are a system of such spirals with the same fixed points and the same constant angle. The lines of level are the orthogonal system of double spirals, with the same fixed points. Figure 32 indicates both systems, one of the systems being shown by dots.

Since the equation of the equiangular spiral is (§ 31)

$$\operatorname{Log} \rho = \alpha\theta + \beta,$$

the equation of the double spiral is

$$\operatorname{Log}(\rho/\rho') = \alpha(\theta - \theta') + \beta,$$

where ρ, θ are measured from one of the fixed points, f, and ρ', θ' from the other fixed point f'. For a system of spirals as in fig. 32, α is a constant but β a parameter.

The substitution is called, in this general case, *loxodromic.*

If we invert the sphere into a plane by using a point of the sphere as centre of inversion, the loxodrome becomes our double spiral; hence the reason for calling the substitution loxodromic. When the centre of inversion is at the pole the loxodrome becomes the single spiral.

Ex. 1. The points of inflexion of all paths in fig. 32 lie on a straight line.

Ex. 2. The velocity in loxodromic motion is proportional to

$$(x-f)(x-f').$$

Ex. 3. In the elliptic substitution the points which move along a straight line have a constant angular velocity about either fixed point.

38. Case of coincident Fixed Points. So far we have considered the fixed points f, f' as distinct; we proceed to the case when they coincide. Since now $f'=f$, the value of k is 1; and the equation

$$\frac{x_1-f}{x_1-f'} = k\frac{x_0-f}{x_0-f'}$$

conveys no information. But let $f'=f+\delta f$, and let δf become 0 in a specified manner, say with an amplitude γ. The above equation is

$$1+\frac{\delta f}{x_1-f'} = \left(1+\frac{c\delta f}{cf+d}\right)\left(1+\frac{\delta f}{x_0-f'}\right),$$

or, dividing by δf and then letting δf tend to zero,

$$\frac{1}{x_1-f} = \frac{1}{x_0-f} + \frac{c}{cf+d}.$$

Map the x-plane on a second plane by writing

$$z = \frac{1}{x-f};$$

we get
$$z_1 = z_0 + \frac{c}{cf+d},$$

a translation of the z-plane.

Hence the lines of flow appropriate to this case of coincident fixed points are a parabolic system of circles which touch at f (§ 23). The lines of level are the orthogonal system, also touching at f.

The substitutions (1), *for which the fixed points coincide, are called parabolic.*

The common tangent to the system of circles is itself a circle of the system, namely a circle through f and ∞. The correspondent x_1 of $x_0=\infty$ lies on this circle. But

$$x_1 - f = (cf+d)/c,$$

or
$$x_1 = 2f + d/c.$$

Hence the common tangent can be constructed by joining f to $2f+d/c$

Ex. The substitution is defined when we assign f and the initial and final values of a point,—say that x_0 is x_0' when x_1 is x_1'. Prove that if we construct a point g so that f and g are harmonic with x_0 and x_1', then x_1 and x_0' are harmonic with f and g. This is then a construction for the point x_1 corresponding to any point x_0.

39. Substitutions of Period Two. Let

$$x_1 = (ax_0+b)/(cx_0+d),\ x_2 = (ax_1+b)/(cx_1+d),$$

and generally $x_{m+1} = (ax_m+b)/(cx_m+d)$; what is the condition that x_n shall coincide with x_0? The question can be answered very readily by taking the equivalent equations

$$z_1 = kz_0,\ z_2 = kz_1 = k^2 z_0, \ldots,\ z_{m+1} = kz_m = \ldots = k^{m+1} z_0;$$

the necessary and sufficient condition is that k^n shall be equal to 1, i.e. k must be an nth root of unity. Since $|k| = 1$, the substitution must be elliptic.

Let us examine the simplest cases $n = 2$ and $n = 3$.

When $n = 2$ we have $k^2 = 1$; and the substitution becomes

$$\frac{x_1-f}{x_1-f'} = \pm\frac{x_0-f}{x_0-f'}.$$

Taking the upper sign, i.e. $k = 1$, the substitution becomes $x = x_0$ and every point remains stationary. When $k = -1$ we have

$$\frac{x_1 - f}{x_1 - f'} + \frac{x_0 - f}{x_0 - f'} = 0,$$

so that (§ 24) x_0 and x_1 are harmonic with f and f'. Combining all the points of the plane into pairs harmonic with two given points f and f we see that the substitution interchanges these points. This arrangement of the points of a plane into pairs is said to be an *involution.*

Any symmetric bilinear substitution between x_0 and x_1, that is any equation of the form

$$cx_0x_1 = a(x_0 + x_1) + b \ldots\ldots\ldots\ldots\ldots\ldots (1),$$

defines an involution. For by a single substitution x_0 becomes x_1 while x_1 becomes x_0; hence by two substitutions x_0 becomes x_0; hence $k = -1$.

The above equation between x_0 and x_1 is the condition that x_0 and x_1 are harmonic with the roots of

$$cx^2 = 2ax + b \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2).$$

Two pairs of points, say p_0, p_1 and q_0, q_1, will determine an involution; for the two equations

$$cp_0p_1 = a(p_0 + p_1) + b,$$
$$cq_0q_1 = a(q_0 + q_1) + b,$$

determine the ratios of a, b, c. The relation (1) is then

$$\begin{vmatrix} x_0x_1, & x_0 + x_1, & 1 \\ p_0p_1, & p_0 + p_1, & 1 \\ q_0q_1, & q_0 + q_1, & 1 \end{vmatrix} = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3),$$

which can be written

$$(x_0 - p_1)(p_0 - q_1)(q_0 - x_1) = (x_0 - q_1)(p_0 - x_1)(q_0 - p_1) \ldots\ldots (4),$$

or in three other equivalent forms obtained by interchanging x_0 and x_1, or p_0 and p_1, or q_0 and q_1.

Ex. Four straight lines can be paired off in three ways. Let the intersections of the pairs be x_0, x_1; p_0, p_1; q_0, q_1. Prove by equating amplitudes and lengths in both sides of (4) that these three pairs of points are in involution.

Geometric Construction for the Partner of $x = \infty$. If we speak of the two points which are paired off as *partners*, then the equation (3) gives the partner of any point x_0 in the involution determined by given pairs. In

particular to find the partner of ∞ we first divide the first row of the determinant by x_0 and then let x_0 tend to ∞. Thus we get

$$\begin{vmatrix} x_1, & 1, & 0 \\ p_0p_1, & p_0+p_1, & 1 \\ q_0q_1, & q_0+q_1, & 1 \end{vmatrix} = 0,$$

or $$x_1(p_0+p_1-q_0-q_1)=p_0p_1-q_0q_1 \qquad (5).$$

Here the origin is unspecified. If we take the origin at q_0 itself the equation is

$$x_1(p_0+p_1-q_1)=p_0p_1,$$

or, if $$p_0+p_1-q_1=q,$$

$$x_1q=p_0p_1 \qquad (6).$$

Since $q_0=0$, this gives the following geometric construction for the partner of ∞:—

Complete the parallelogram $p_0q_1p_1q$ and make the triangle $p_0q_0x_1$ similar to the triangle qq_0p_1. Then x_1 is the point required.

This point is called the *centre of the involution.*

The Double Points of the Involution. The points which are their own partners are the fixed points of the substitution, but with reference to the involution they are usually called the double points; for in the pairing off of the points of the plane each is counted twice. To obtain them when two pairs are given, we have, by writing $x_1=x_0=x$ in (3) or (4), a quadratic in x. But this does not suggest at once a geometric construction.

We can construct them very easily from knowing the centre; for if we write $x_1=0$ in (5) and thus take the centre as origin, the condition that p_0 and p_1, q_0 and q_1, 0 and ∞ are in involution is, as in (6),

$$p_0p_1=q_0q_1;$$

whence the double points are the square roots of p_0p_1, and lie on a straight line through the origin which bisects the angle p_00p_1. And now if we state this in a covariant manner we can dispense with first finding the centre. For the straight line which bisects p_00p_1 is a circle through 0 and ∞ which makes opposite angles with the arcs $0p_0\infty$ and $0p_1\infty$.

Thus, generally, if p_0p_1 and q_0q_1 are the given pairs, draw the arcs $q_0p_0q_1$ and $q_0p_1q_1$, and draw a circle through q_0 and q_1 making opposite angles with these arcs. This circle contains the double points. And interchanging the pairs we obtain another such circle, which cuts the former at the points sought.

This construction fails, however, when the given pairs lie on a circle and are not interlaced. For then the two circles of the construction coincide with the circle on which the pairs lie. To meet this case we can apply the fact that if two circles A, B intersect, and A' be orthogonal to both, A' intersects either circle at points harmonic with the intersections of A and B. If then A' and B' are circles orthogonal with both A and B, each marks off on A points harmonic with the intersections of A and B. Thus if p_0, p_1 and q_0, q_1 are on a circle or straight line A, and are not interlaced, we draw through p_0, p_1 and q_0, q_1 circles A' and B' orthogonal to A, and construct any other

circle B orthogonal to both A' and B'. The points where B intersects A are the sought double points.

Ex. 1. When p_0, p_1 and q_0, q_1 are on a circle, and the straight lines p_0p_1 and q_0q_1 meet at a point r outside the circle, prove that the points of contact of the tangents from r are the double points of the involution determined by p_0, p_1 and q_0, q_1. Prove that the partner of r in this involution is the centre of the circle.

Ex. 2. The above general construction for the double points is by means of angles. Discuss the correlative construction by means of lengths.

40. Reduction of Four Points to a Canonic Form. It is often convenient, in discussing involutions, to suppose the double points to be 0 and ∞. The bilinear relation is then merely $z_0 + z_1 = 0$, and two pairs are opposite corners of a parallelogram. In mapping two given pairs of points into this parallelogram form, we may further suppose that one of the points maps into 1; the equations which give the pairs of points are thus brought by a bilinear transformation into the forms

$$z^2 = 1, \quad z^2 = z_0^2.$$

To determine z_0, we use the principle that a double ratio is unchanged by a bilinear transformation. The double ratio

$$(z_0, \ -z_0, \ 1, \ -1)$$

is

$$\frac{z_0 - 1}{z_0 + 1} \cdot \frac{-z_0 + 1}{-z_0 - 1}, \text{ or } \left(\frac{z_0 - 1}{z_0 + 1}\right)^2.$$

Hence if p_0, p_1 and q_0, q_1 be the given pairs we can take

$$\left(\frac{z_0 - 1}{z_0 + 1}\right)^2 = (p_0, q_0; p_1, q_1),$$

whence we have for z_0 two reciprocal values, either of which can be taken.

It will be observed that it is not necessary to calculate the double points.

Ex. 1. Prove that the equation $(x^2 - 1)(x^2 - 2x + 2) = 0$ can be bilinearly transformed into $(z^2 - 1)[z^2 + (2 \pm \sqrt{5})^2] = 0$.

Ex. 2. When two pairs form a parallelogram, verify the construction for the double points given in the preceding article.

Ex. 3. Four points can be paired off in three ways and therefore determine three involutions. Prove, by considering the parallelogram, that the three pairs of double points are mutually harmonic, and that the

partners of any point in two of the involutions are themselves partners in the third.

41. Substitutions of Period Three. The case $n=3$ of § 39, that is the case in which the substitution, when applied three times, brings each point back to its initial position, arises

when $$k^3 = 1,$$

whence $$k^2+k+1=0,\ k=\exp(\pm 2\pi i/3).$$

Taking three arbitrary points x_0, x_1, x_2, we can of course determine a substitution which will bring them into the positions x_1, x_2, x_0; we now see that this substitution is of the

form $$(x_1-f)/(x_1-f')=\exp(\pm 2\pi i/3)\,.\,(x_0-f)/(x_0-f'),$$

where f and f' are to be determined.

This way of mapping a triad of points into itself is very closely connected with the solution of a cubic equation. For the essential point is that by writing $(ax+b)/(cx+d)=z$ we can map the triad into any assigned regular triad in the z-plane; say into 1, v, v^2.

Thus the cubic whose roots are x_0, x_1, x_2, can be written in the form

$$z^3=1,$$

by writing $$(ax+b)/(cx+d)=z.$$

That is, the cubic itself can be written

$$(ax+b)^3=(cx+d)^3,$$

or say $$(x-f)^3=\lambda(x-f')^3.$$

Suppose the cubic to be $x^3+3\alpha x+\beta=0$; then we have, comparing coefficients,

$$f=\lambda f',\ f^2-\lambda f'^2=\alpha(1-\lambda),\ f^3-\lambda f'^3=-\beta(1-\lambda);$$

whence $$ff'=-\alpha,\ ff'(f+f')=-\beta,$$

so that f and f' are determined by solving a quadratic.

We have then

$$(x-f)/(x-f')=\sqrt[3]{\lambda}=\sqrt[3]{f/f'},$$

the three cube roots giving the three roots of the cubic.

Ex. 1. The points x_0, x_1, x_2 being mapped into 1, v, v^2, the points f, f' are at the same time mapped into 0, ∞. Hence corresponding double ratios of x_0, x_1, x_2, f and 1, v, v^2, 0 are equal. Hence prove that

$$f' = -\frac{x_1 x_2 + v x_2 x_0 + v^2 x_0 x_1}{x_0 + v x_1 + v^2 x_2},$$

$$f = -\frac{x_1 x_2 + v^2 x_2 x_0 + v x_0 x_1}{x_0 + v^2 x_1 + v x_2}.$$

Ex. 2. Prove that the angle which the stroke from f to f' subtends at any of the points $x_0 x_1 x_2$ is a third of the angle which it subtends at their centroid.

Ex. 3. A bilinear transformation can be found which will map a, b, c, d into b, c, d, a only when a, c and b, d are harmonic pairs[1].

[1] The periodic $n-ad$, or figure formed by a point which returns to its original position after n substitutions, can be regarded as the map of a regular $n-ad$,—that is of the vertices of a regular polygon. This is not the place to enter into geometric considerations for their own sake, but as the figure in question has received some attention from geometers (see for example Casey's *Sequel to Euclid* and *Analytic Geometry*), it is proper to point out that it can be obtained from the regular $n-ad$ by a single inversion with regard to an arbitrary point of space; and that therefore it is a special projection of the regular $n-ad$.

CHAPTER VI.

LIMITS AND CONTINUITY.

42. Concept of a Limit. The discussion of irrational numbers brings us face to face with difficulties which cluster round the notions of continuity, limit, and convergence. As these notions are woven into the very texture of our subject, it will not be amiss here to indicate to the reader some of their aspects which may have escaped his attention when studying algebra and the differential calculus. For simplicity we will take the variable real.

The sequence $4, 3\frac{1}{2}, 3\frac{1}{3}, 3\frac{1}{4}, \ldots$ suggests at once the number 3. This number 3 is not itself a member of the sequence, but the numbers of the sequence tend to it (or converge to it) and *can be made to differ from it by as little as we please.* We are therefore justified in saying that the function $\xi + 3$ tends to the limit 3, when ξ passes through the values $1, 1/2, 1/3, \ldots$. This statement is expressed symbolically in the form $\lim_{\xi=0} (\xi + 3) = 3$. Observe that ξ did not take the value 0, for 0 is not one of the set $1, 1/2, 1/3, \ldots$.

Let us now state the matter generally. The numbers $\xi_1, \xi_2, \xi_3, \ldots$ of a sequence are said to *tend to the limit* α, when to every positive number ϵ there corresponds a positive integer μ* such that for ξ_μ and for all later members ξ_n of the sequence we have

$$|\xi_n - \alpha| < \epsilon.$$

* The number μ will be different in general for different ϵ's. The number ϵ is said to be arbitrarily small and given in advance. It may be supposed assigned by an

When, with the same notation, we have $\xi_n > 1/\epsilon$ (or $< -1/\epsilon$) we say that the numbers ξ_n *tend to the limit* $+\infty$ (*or* $-\infty$).

43. Distinction between "value when $\xi=\alpha$" and "limit when $\xi=\alpha$." To return to the special example, we have seen that ξ did not take the value $\xi=0$; suppose now that we assign to ξ the value 0, we have at once the value 3 for the function $\xi+3$. We express this in the following way:

$$\underset{\xi=0}{\text{value}}\,(\xi+3)=3.$$

In this example

$$\lim_{\xi=0}(\xi+3)=\underset{\xi=0}{\text{value}}\,(\xi+3).$$

It is most important that the reader should appreciate that it is far from being true universally that the limit of an expression when ξ tends to α is the value of the expression when ξ is α.

We will consider a case in which the equality breaks down.

Suppose that $f\xi=\dfrac{\xi^2-1}{\xi-1}$ and that $\alpha=1$. To fix ideas let ξ approach 1 through the sequence of values $1+1$, $1+1/2$, $1+1/3,\ldots$, then $\dfrac{\xi^2-1}{\xi-1}$ (or $\xi+1$) takes the values $3, 2\frac{1}{2}, 2\frac{1}{3},\ldots$. The left-hand side of the equation

$$\lim_{\xi=1}\frac{\xi^2-1}{\xi-1}=\underset{\xi=1}{\text{value}}\,\frac{\xi^2-1}{\xi-1}$$

has therefore the definite value 2; but the right-hand side is meaningless and the equation breaks down.

44. Upper and Lower Limits. The numbers of the sequence $4, 3\frac{1}{2}, 3\frac{1}{3}, 3\frac{1}{4},\ldots$ lie between 0 and 10; this interval can be contracted to the narrower interval 2 to 6, and this again can be contracted. Suppose that we take the narrowest interval which contains all the numbers; this is evidently the interval from 3 to 4. The number 3 is called the *lower limit* and the number 4 is called the *upper limit* of the sequence.

opponent. The opponent names as many numbers ϵ as he pleases, and one has to assign for each a suitable μ. Of course an algebraic inequality, such as $\mu > 1/\epsilon$, is the usual means of establishing the criterion. In arguments about limits ϵ is restricted to the above meaning, with the consequent advantage that the words 'arbitrarily small' and 'given in advance' can be omitted.

It is clear that the notions upper and lower limit coincide with the notions of maximum and minimum, whenever the sequence contains a maximum number and a minimum number. For example

(i) the sequence $4, 3\frac{1}{2}, 3\frac{1}{3}, 3\frac{1}{4}, \ldots$ has an upper limit 4 which is also a maximum: there is a lower limit 3, but no minimum;

(ii) the sequence $3, 4, 3\frac{1}{2}, 3\frac{1}{3}, 3\frac{1}{4}, \ldots$ has an upper limit 4 which is a maximum and a lower limit 3 which is a minimum;

(iii) the sequence formed from the numbers $3 + 1/n$, $4 - 1/n$ by giving to n the values $2, 3, 4, \ldots$, has the upper limit 4 and the lower limit 3, but it has neither a maximum nor a minimum.

45. Every Sequence of constantly increasing Real Numbers admits a Finite or Infinite Limit. Let the real variable ξ increase constantly; either there is no real number α which is greater than all the values assumed by ξ, or there is such a number. In the former case ξ can become indefinitely large; in the latter case there is a number γ to which ξ comes arbitrarily close, or to which ξ attains, but beyond which ξ does not pass. In the one case we say that the limit of ξ is $+\infty$, while in the other the limit of ξ is γ. In the latter case the limit is also the upper limit (§ 44); in the former case we shall say that $+\infty$ is the upper limit.

Similarly when ξ decreases constantly it tends to the limit $-\infty$ or else to a finite limit γ'. In the former case the lower limit is $-\infty$, in the latter γ'.

The existence of the numbers γ, γ' is not a new assumption; it is a consequence of the arithmetic definition of irrational numbers. For all real numbers can be separated into the two classes,

(1) the numbers which are greater than all the values that ξ is allowed to take;

(2) the numbers which are exceeded by some of the values of ξ.

From these two classes there is no difficulty in picking out two sequences which will define the number γ; and similar reasoning establishes the existence of γ'.

46. Every Sequence of Real Numbers has an Upper and a Lower Limit.

First let there be an integer α which is less than all the numbers of the sequence and construct the sequence

$$\alpha, \alpha+1, \alpha+2, \ldots\ldots.$$

Passing from left to right along this sequence let (α_1, α_1+1) be the first interval which contains a number ξ. Divide this interval into tenths and consider the new sequence

$$\alpha_1, \alpha_1+1/10, \alpha_1+2/10, \ldots, \alpha_1+9/10, \alpha_1+1.$$

Passing again from left to right let $(\alpha_2, \alpha_2+1/10)$ be the first interval which contains a number ξ. Divide this interval into tenths and proceed as before. In this way we construct

$$\alpha_1, \alpha_2, \ldots\ldots\ldots, \alpha_m, \ldots\ldots\ldots,$$
$$\alpha_1+1, \alpha_2+1/10, \ldots\ldots\ldots, \alpha_m+1/10^{m-1}, \ldots\ldots\ldots,$$

which are sequences of the kind considered in ch. I. (§ 6). They define therefore a rational or irrational number γ'; we shall prove that this number is the lower limit, that is that there is no number ξ which is less than γ' and that there are numbers ξ which differ from γ' by less than any assigned positive number however small. The latter part of this statement is true by reason of the presence of at least one number ξ in every interval $(\alpha_m, \alpha_m+1/10^{m-1})$, no matter how large m may be. To prove the former part of the statement, assume that there is a ξ, say ξ_r, which is less than γ'. Then because $\alpha_1, \alpha_2, \alpha_3, \ldots\ldots$ tend to the limit γ', it must be possible to find an α, say α_s, which is nearer than ξ_r to γ'; thus we have a ξ which is less than an α, contrary to supposition. It follows that γ' is the lower limit.

What has been proved may be formulated as follows:—

Given that all the values of a real variable ξ are greater than a finite integer α, there exists one and only one finite number γ' with these properties:—(1) *no value of ξ is less than γ'*, (2) *at least one value of ξ is less than $\gamma'+\epsilon$, where ϵ is an arbitrarily small positive number assigned in advance. This number γ' is the lower limit of the values of ξ.*

Secondly, when there is no integer a which is less than all the numbers of the sequence, we say, as in § 45, that $-\infty$ is the lower limit. Thus in all cases there is a lower limit.

The corresponding theorem for the upper limit can be proved in the same way.

47. The necessary and sufficient condition that a Sequence tends to a finite Limit. When we have reason to suspect that the numbers ξ_n tend to a limit α, we can verify whether this is so or not by using the definition of § 42, namely

$$|\xi_n - \alpha| < \epsilon, \qquad (n \geqq \mu).$$

But it is often not possible to write down the value of α and it is therefore essential to have an equivalent test which will not presuppose a knowledge of α. *The new condition is that it shall be possible to find among the ξ's a number ξ_μ such that the absolute value of the difference of any two numbers of the sequence ξ_μ, $\xi_{\mu+1}$, $\xi_{\mu+2}$,... shall be less than an arbitrarily small positive number ϵ given in advance.*

(A) The condition is necessary; for let α be the limit and let μ be so chosen that

$$|\alpha - \xi_n| < \epsilon/2, \ |\alpha - \xi_{n'}| < \epsilon/2,$$

where $n' > n$ and $n \geqq \mu$. Then

$$|\xi_{n'} - \xi_n| \leqq |\alpha - \xi_n| + |\alpha - \xi_{n'}| < \epsilon.$$

(B) The condition is sufficient. The spirit of the proof can be appreciated best when use is made of geometric considerations, though the proof itself is strictly arithmetic. Represent ξ_1, ξ_2, ξ_3,... by points on the axis of ξ. In virtue of the condition imposed upon the ξ's we can say that from a certain point ξ_μ onwards all the points $\xi_{\mu+1}$, $\xi_{\mu+2}$,...... lie within a distance ϵ of ξ_μ on either side; so that if $\xi_\mu - p = q - \xi_\mu = \epsilon$, every number

Fig. 33.

$\xi_{\mu+1}$, $\xi_{\mu+2}$,... lies within the interval pq; in particular $\xi_{\mu'}$ lies

between p and q, if $\mu' > \mu$ (fig. 33). By choosing μ' sufficiently large, we can affirm that every point $\xi_{\mu'+1}$, $\xi_{\mu'+2}$, $\xi_{\mu'+3}, \ldots$ lies within a new interval $p'q'$, where $\xi_{\mu'} - p' = q' - \xi_{\mu'} = \epsilon/2$. This interval $p'q'$ may extend beyond p or q but we can cast away the projecting part since all the points ξ_μ, $\xi_{\mu+1}, \ldots$ lie in pq. Attending then only to the interval common to pq and $p'q'$ we observe that its length is at most half the length of pq, and that it contains all the points $\xi_{\mu'}$, $\xi_{\mu'+1}, \ldots$. Among the numbers $\mu'+1$, $\mu'+2$, $\mu'+3, \ldots$ we can find a number μ'' sufficiently large to secure that all the points $\xi_{\mu''}$, $\xi_{\mu''+1}$, $\xi_{\mu''+2}, \ldots$ shall lie within an interval whose length is at most half that of $p'q'$, this interval lying wholly within each of the preceding intervals. By proceeding in this way we can get a succession of intervals whose lengths tend to zero; furthermore the left-hand and right-hand extremities of these intervals, when connected with the origin by strokes, give two sequences of strokes in ascending and descending order respectively. Hence the conditions of § 7 are fulfilled and the strokes define a point α, which is the limit whose existence we sought to establish.

The non-terminating decimal affords a good illustration.. When we take the first μ digits after the decimal point, let the number obtained be ξ_μ. Then $\xi_{n'} - \xi_n < 1/10^\mu$, and this can be made less than any assigned ϵ by taking μ large enough. Therefore there is a limit of the sequence ξ_1, ξ_2, $\xi_3, \ldots$. If the decimal terminates, the limit is attained

A sequence which tends to a *finite* limit is said to be *convergent.*

48. Real Functions of a Real Variable. The most general definition of a real function of a real variable is as follows:—

When one and only one real value η is assigned to each value of a real variable ξ, η is said to be a one-valued function of ξ. When two or more real values are assigned to each ξ, η is said to be a many-valued function of ξ.

We restrict ourselves in this chapter to one-valued functions.

This definition admits functions which cannot be represented graphically. Let, for example, η be equal to 0 for all rational values of the real variable ξ and to 1 for all irrational values of that variable; then the points (ξ, η) trace out no plane curve in the sense ordinarily accepted for the term 'curve.'

Similarly in the case of two or more independent *real* variables, what is implied by the statement that a dependent variable is a function of these independent variables is conveyed adequately by the adjective 'dependent'; the former is to be given when the latter are given. We lay emphasis on this; for we propose to consider these general functions in the case where the variable or variables are *real*, although the only functions of the *complex* variable x that will be discussed in this work will be drawn from a very general, though not completely general, class of functions known as 'analytic.'

To keep the two uses of the word 'function' as distinct as possible we shall use $f\xi$, $f(\xi, \eta)$, when simple dependence on real variables is meant; whereas fx, where $x = \xi + i\eta$, will indeed denote dependence on x,—therefore also on ξ, η,—but dependence of that special kind which will be expressed later by *analytic function*.

49. Continuity of a Function of a Real Variable. Suppose that when the real variable ξ tends to the limit α by increasing values of ξ, the function $f\xi$ tends to a limit; we can denote this limit by $\lim\limits_{\alpha - 0} f\xi$; similarly we use $\lim\limits_{\alpha + 0} f\xi$ to denote the limit (also assumed to exist) for $f\xi$ when ξ tends to α by decreasing values. With this notation we can define continuity as follows:

Let $f\xi$ be a one-valued function of the real variable ξ, which is defined for all values of an interval $\alpha - \delta$ to $\alpha + \delta$, then $f\xi$ is said to be continuous at α, if

$$\lim_{\alpha - 0} f\xi = \lim_{\alpha + 0} f\xi = \operatorname*{value}_{\xi = \alpha} f\xi.$$

The function is discontinuous when any one of the three numbers in these equations is non-existent, and when the three numbers exist but are not all equal. For example $\dfrac{\xi^2 - \alpha^2}{\xi - \alpha}$ is not continuous at $\xi = \alpha$, because the expression has no meaning when ξ is equal to α; here $\lim\limits_{\alpha - 0} \dfrac{\xi^2 - \alpha^2}{\xi - \alpha} = \lim\limits_{\alpha + 0} \dfrac{\xi^2 - \alpha^2}{\xi - \alpha} = 2\alpha$, but $\operatorname*{value}\limits_{\xi = \alpha} \dfrac{\xi^2 - \alpha^2}{\xi - \alpha}$ is non-existent.

Suppose that when $\xi < 1$ we have $\eta = f\xi = 1$, and when $\xi \geqq 1$

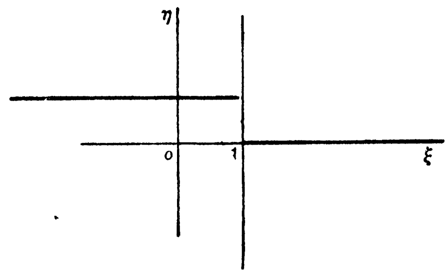


Fig. 34.

we have $\eta = f\xi = 0$ (fig. 34); then $\lim\limits_{1-0} f\xi = 1$, whereas $\lim\limits_{1+0} f\xi = 0$, and $\underset{\xi=1}{\text{value}} f\xi = 0$. Here again there is discontinuity.

The necessary and sufficient condition for the continuity of $f\xi$ at $\xi = \alpha$, where $f\xi$ is defined for $\xi = \alpha$, may also be stated in the following way:—

The function $f\xi$ is continuous at $\xi = \alpha$, when there exists an interval of the axis of ξ,—say the interval from $\alpha - \delta$ to $\alpha + \delta$, where δ is real,—such that at every point ξ of this interval, we have

$$|f\xi - f\alpha| < \epsilon,$$

where ϵ is an arbitrarily small positive number which is chosen in advance. It is implied here that the function is defined when $\xi = \alpha$.

50. A Continuous Function of a Real Variable attains its Upper and Lower Limits. Let $f\xi$ be continuous at all points ξ for which $\alpha \leqq \xi \leqq \beta$, α and β being finite. This interval may be called the *closed interval* (α, β); closed, because the interval contains both extremities α and β. Within the closed interval (α, β) the values $f\xi$ have an upper limit γ and a lower limit γ'. We wish to prove the important theorem that there exists in the interval a point at which $f\xi$ is *exactly* γ, and a point at which $f\xi$ is *exactly* γ'. In other words we wish to show that the values of the continuous function $f\xi$ admit a maximum and a minimum value, when ξ takes all values within the closed

interval. The following lemma is required in the proof of this theorem :—

Let one of the closed intervals $(\xi_0 - \delta,\ \xi_0 + \delta)$, $(\xi_0,\ \xi_0 + \delta)$, $(\xi_0 - \delta,\ \xi_0)$, *where* δ *is an arbitrarily small positive number, be associated with the point* ξ_0 *of the closed interval* (α, β), *the first, second, or third being chosen according as* ξ_0 *is interior to or at the extremities* α, β *of the interval* (α, β); *also let these intervals lie within* (α, β). *Then we can assert that there is at least one point* ξ_0 *of the interval* (α, β) *such that the upper limit of* $f\xi$ *for the associated interval is exactly* γ, *no matter how small* δ *may be taken.*

To establish the truth of this lemma we must prove these two propositions :—

(1) *Given that* (α', β') *is contained within* (α, β), *then the upper limit of* $f\xi$ *for* (α', β') *is at most* γ;

(2) *Given that* (α, β) *is divided into two equal or unequal parts, then the upper limit of* $f\xi$ *for at least one of these parts is equal to* γ.

As regards (1) observe that if the upper limit of $f\xi$ for (α', β') be $\gamma + \gamma_1$ where γ_1 is a positive number, then there are values of $f\xi$ which exceed any assigned number which is less than $\gamma + \gamma_1$, and therefore exceed the number γ. This is contrary to supposition ; for there are by supposition no values of $f\xi$ in (α, β), much less in (α', β'), which exceed γ. If the upper limit of $f\xi$ for (α', β') be $+\infty$, the upper limit of $f\xi$ for (α, β) must also be $+\infty$.

As regards (2) we can use a very similar argument. When γ is finite, the upper limits for the two parts may be equal, in which case both must be equal to γ, or they may be unequal, in which case the greater of the two must be equal to γ. When, on the other hand, γ is $+\infty$ we must have at least one of the two new upper limits equal to $+\infty$.

We can now proceed to the proof of the lemma. Divide (α, β) into two equal parts ; for one at least of these, say (α_1, β_1), the corresponding upper limit is γ. Divide (α_1, β_1) into two equal parts and select a half, say (α_2, β_2), which yields an upper

limit γ. By continuing this subdivision, we are led to two sequences

$$\alpha.\ \alpha_1, \alpha_2, \alpha_3, \ldots\ldots, \alpha_n, \ldots\ldots,$$

$$\beta, \beta_1, \beta_2, \beta_3, \ldots\ldots, \beta_n, \ldots\ldots,$$

which are composed respectively of ascending and descending numbers. These sequences define a rational or irrational number ξ_0, since $\alpha_n - \beta_n, = (\alpha - \beta)/2^n$, tends to zero when n tends to infinity; this number ξ_0 lies within (α_n, β_n).

Within (α_n, β_n) the upper limit of $f\xi$ is γ; hence the upper limit of $f\xi$ in any interval $(\xi_0 - \delta, \xi_0 + \delta)$ which contains (α_n, β_n) is at least γ; since (α_n, β_n) shrinks indefinitely as n grows, we can make the interval $(\xi_0 - \delta, \xi_0 + \delta)$ as small as we please. Again since $(\xi_0 - \delta, \xi_0 + \delta)$ is merely a part of (α, β) we can assert at once that the upper limit of $f\xi$ in the former interval is at most equal to the upper limit, γ, of $f\xi$ in the latter interval. Since the upper limit of $f\xi$ for the interval $(\xi_0 - \delta, \xi_0 + \delta)$ is at least equal to and at most equal to γ, it must be exactly equal to γ. This is on the assumption that γ is finite: when γ is $+\infty$ the upper limit of $f\xi$ for the interval $(\xi_0 - \delta, \xi_0 + \delta)$ is not less than γ and is therefore $+\infty$. This completes the proof of the lemma. It is evident that the lemma is still true when the words 'upper limit' are replaced by 'lower limit,' and γ by γ'.

We have spoken of $\gamma = +\infty$ as if it were a possibility, but the lemma shows that γ cannot be $+\infty$ when $f\xi$ is continuous. For

(a) the supposition $\gamma = +\infty$ implies, by the lemma, that the values of $f\xi$ increase indefinitely as ξ tends to ξ_0;

(b) the supposition that $f\xi$ is continuous at ξ_0 implies that the values of $f\xi$ differ very little from the finite value $f\xi_0$ when ξ approaches very near ξ_0;

and the statements (a), (b) are irreconcilable.

We can now prove the theorem that *a continuous function $f\xi$ attains each of its two limits at least once.*

By reason of the continuity of $f\xi$ at ξ_0, there exists an interval $(\xi_0 - \delta, \xi_0 + \delta)$ such that for every point ξ of this interval we have

$$|f\xi - f\xi_0| < \epsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1),$$

where ϵ is an arbitrarily small positive number assigned in advance. And by reason of the lemma we can assert that there *may* exist within $(\xi_0 - \delta,\ \xi_0 + \delta)$ a point ξ such that $f\xi = \gamma$, and that there *must* exist a point ξ such that

$$f\xi > \gamma - \epsilon.$$

Thus, for the same values of ξ as in (1), we have

$$|f\xi - \gamma| < \epsilon \qquad\qquad (2).$$

Hence, combining (1) and (2), we have

$$|f\xi_0 - \gamma| < 2\epsilon \qquad\qquad (3).$$

But $f\xi_0$, γ are *fixed* quantities, and therefore the absolute value of their difference is a fixed quantity. It is impossible that a *definite* positive number however small should always satisfy the inequality (3), unless this definite number be zero, for ϵ can be made as small as we please. Hence

$$f\xi_0 - \gamma = 0;$$

a result which proves the theorem, so far as the upper limit is concerned. The proof for the lower limit is precisely analogous.

51. Functions of two independent Real Variables. Suppose that a real value ζ is assigned to each point (ξ, η), or $\xi + i\eta$, in a certain region Γ of the plane,—for example for the interior of, and on, a rectangle or circle,—then ζ is for the region in question *a real function of the two real variables* ξ, η.

(I) *The values* ζ *admit an upper and a lower limit.* This is an immediate consequence of what we have said (in § 46) on the existence of an upper and lower limit for a single real variable; in this case the variable is ζ.

This upper limit can be called γ, where γ is finite or $+\infty$, and the lower limit γ', where γ' is finite or $-\infty$.

(II) Surround the region Γ by a rectangle $a_1a_2a_3a_4$ whose sides are parallel to the axes of ξ and η. Through the centre b_1 of the rectangle draw lines parallel to the sides, thus creating four new rectangles. Of these new rectangles consider only such as contain part of Γ. In at least one of these rectangles, say $b_1b_2b_3b_4$ $(b_3 = a_3)$, the upper limit of the values of ζ for the

points common to Γ (the contour included) and to the rectangle is exactly γ, the upper limit of ζ for the whole of Γ. Divide this

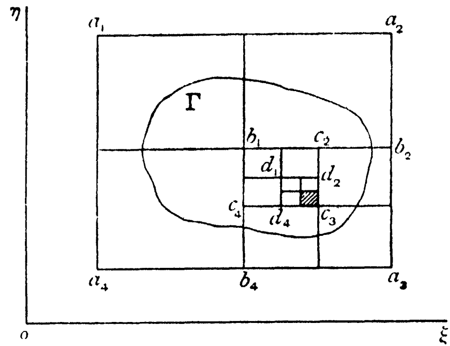


Fig. 35.

rectangle into four equal parts as before, and select a part which yields the upper limit γ for the associated values of ζ; let this part be $c_1c_2c_3c_4$ ($c_1 = b_1$), and proceed as before. Then we get a sequence of rectangles

$$a_1a_2a_3a_4,\ b_1b_2b_3b_4,\ c_1c_2c_3c_4,\ d_1d_2d_3d_4, \ldots,$$

each of which contains all that follow; the areas of these rectangles tend to zero and there is one and only one point (ξ_0, η_0) which lies within all the rectangles of the sequence. This assumption is of the same nature as the assumption made in § 7. For each rectangle provides two abscissae and two ordinates; for example $c_1c_2c_3c_4$ provides the abscissae of c_1 (or c_4) and c_2 (or c_3) and the ordinates of c_3 (or c_4) and c_1 (or c_2). And hence the rectangles lead to ascending and descending sequences of abscissae which converge to an abscissa ξ_0 and to ascending and descending sequences of ordinates which converge to an ordinate η_0. The combination of the sequences leads to a point (ξ_0, η_0).

It is proper to define here the expression *neighbourhood of* (ξ, η). All that is meant by this is a region that consists of points on and within some circle whose centre is (ξ, η), and whose radius is not zero. We can see, precisely as in § 50, that :—

The values of the real function ζ of the two real variables ξ, η for the region Γ of the plane admit the same upper limit for every

neighbourhood, contained within Γ, *of the above point* (ξ_0, η_0) *as for the whole region* Γ.

It is almost unnecessary to say that when a neighbourhood of (ξ_0, η_0) lies partially outside Γ,—this will happen for example when (ξ_0, η_0) is on the boundary of Γ,—the theorem will still apply, provided that account be taken only of such points of the neighbourhood in question as lie within or on the boundary of Γ.

(III) **Continuity of** $f(\xi, \eta)$. Suppose that $f(\xi, \eta)$ is defined for all points of a region Γ which covers a continuous portion of the plane; for instance Γ may consist of all points on and interior to the ellipse $\frac{\xi^2}{\alpha^2} + \frac{\eta^2}{\beta^2} = 1$. We shall say that *the function* $f(\xi, \eta)$ *is continuous at a point* (ξ_0, η_0) *in the interior of* Γ, *when there can be found a neighbourhood of the point such that for every point* (ξ, η) *of this neighbourhood, we have*

$$|f(\xi, \eta) - f(\xi_0, \eta_0)| < \epsilon,$$

where ϵ *is an arbitrarily small positive number given in advance.*

52. A Continuous Function $f(\xi, \eta)$ attains its Upper and Lower Limits. It is now possible to prove the following important theorem on the upper and lower limits of a continuous function:—

There is at least one point of Γ *at which the value of* $f(\xi, \eta)$ *is exactly* γ *and at least one point at which the value is exactly* γ'.

Let (ξ_0, η_0) be found by means of the sequence of rectangles as explained above. Then $f(\xi_0, \eta_0)$ must be equal to γ; for if possible let it be equal to $\gamma - \delta$, where δ is a positive number.

(*a*) Because $f(\xi, \eta)$ is continuous, it follows that there is a neighbourhood of (ξ_0, η_0) for every point (ξ, η) of which

$$|f(\xi, \eta) - f(\xi_0, \eta_0)| < \delta/2.$$

Hence for this neighbourhood $f(\xi, \eta)$ cannot be greater than $f(\xi_0, \eta_0) + \delta/2$, that is than $\gamma - \delta/2$.

(*b*) Because $f(\xi, \eta)$ admits the upper limit γ within every

neighbourhood of (ξ_0, η_0), there must be a point (ξ, η) in the particular neighbourhood considered in (*a*), for which

$$f(\xi, \eta) > \gamma - \delta/2,$$

for the values of $f(\xi, \eta)$ must approach arbitrarily close to γ and some of them must therefore exceed $\gamma - \delta/2$.

As (*a*) and (*b*) contradict each other, it follows that $f(\xi_0, \eta_0) = \gamma$. Similarly it can be shown that $f(\xi, \eta)$ attains the value γ'.

53. Uniform Continuity of a Function of one Real Variable. Suppose that a real function $f\xi$ of a real variable ξ is continuous at every point ξ of a closed interval (α, β) given by $\alpha \leqq \xi \leqq \beta$. Because of the continuity of $f\xi$ there exists for every point ξ between α and β an interval $(\xi - h, \xi + h)$ such that for every pair of values ξ', ξ'' included in (α, β) and in $(\xi - h, \xi + h)$, we have

$$|f\xi' - f\xi''| < \epsilon \text{ ....} (1).$$

Let us understand by the *oscillation* of $f\xi$ for values ξ in the interval $(\xi - h, \xi + h)$ the difference between the lowest value that $f\xi$ cannot exceed and the highest value below which $f\xi$ cannot fall; then the condition (1) can be replaced by the equally serviceable condition that there shall exist an interval $(\xi - h, \xi + h)$ such that the oscillation of $f\xi$ in this interval is less than ϵ. There is still complete indeterminateness for the value to be assigned to h; for if any value h_1 of h satisfy the requirements, so do all values of h between 0 and h_1. The infinitely many values of h that are associated with an assigned ξ have an upper limit ρ, whose value depends on ξ. Let us use the interval $(\xi - \rho, \xi + \rho)$ in place of $(\xi - h, \xi + h)$. Before going further one remark must be made to clear away possible misconceptions: at points ξ near the extremities α, β of (α, β) it may happen that part of the interval $(\xi - \rho, \xi + \rho)$ lies outside (α, β), and whenever this happens this part is to be thrown aside.

The breadth 2ρ of the interval at ξ will vary as ξ goes from α to β (fig. 36), and must have a lower limit Δ.

With a view to the *a priori* possibility that Δ may prove to be zero, continuity has been classified as

(i) *uniform continuity*, when $\Delta > 0$;

(ii) *non-uniform continuity*, when $\Delta = 0$.

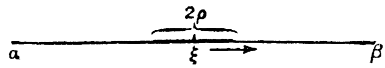


Fig. 36.

Suppose that $\Delta > 0$; the essential fact to grasp in this connexion is that one and the same value of h, namely $h = \Delta$, will serve for the inequality (1), *whatever be the position of ξ in the interval* (α, β). The following theorem disposes of the possibility of the continuity being non-uniform:—

If a real function $f\xi$ of a real variable ξ be continuous in a closed interval (α, β), *the continuity must be uniform.*

As soon as it is shown that ρ is a continuous function of ξ, the theorem follows very readily. For in this case ρ attains its lower limit Δ at some point c of (α, β), by § 50; and $\Delta = 0$ means that there is *no* interval at c within which the oscillation is less than ϵ, contrary to supposition.

Let ξ_1 be a point of the interval $(\xi - \rho, \xi + \rho)$, and let ρ_1 denote $\rho(\xi_1)$. Evidently the interval $(\xi_1 - \rho_1, \xi_1 + \rho_1)$ that belongs to ξ_1 must extend at least to the nearer and at most

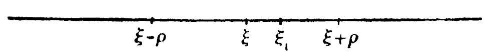


Fig. 37.

to the farther of the two points $\xi - \rho$, $\xi + \rho$. In fig. 37 we have drawn ξ_1 a little to the right of ξ, and in this case

$$-(\xi_1 - \xi) \leqq \rho_1 - \rho \leqq \xi_1 - \xi.$$

In any case, when ξ_1 is near ξ,

$$|\rho - \rho_1| \leqq |\xi - \xi_1|;$$

and this inequality implies the continuity of ρ at ξ, which is what we had to prove.

54. Uniform Continuity of a Function of two Real Variables. Let $f\xi$ be replaced by $f(\xi, \eta)$, where ξ and η are real; and let the interval of continuity (a, b) be replaced by a region of continuity, Γ, the boundary of Γ

included. Here, of course, (ξ, η) is treated as a point referred to rectangular coordinate axes. Finally let $f(\xi, \eta)$ be real, one-valued, and continuous at each point (ξ_0, η_0) of Γ.

To simplify matters let (ξ_0, η_0) lie in the interior of Γ. The continuity of $f(\xi, \eta)$ at (ξ_0, η_0) is expressed in the statement that the inequality

$$|f(\xi, \eta)-f(\xi_0, \eta_0)| < \epsilon,$$

is satisfied by all points within a circle of radius ρ' whose centre is (ξ_0, η_0). The upper limit ρ, or $\rho(\xi_0, \eta_0)$, of the values ρ' can be used in place of ρ'. To prove that the continuity of $f(\xi, \eta)$ in Γ is uniform it is necessary to show that the lower limit of the values $\rho(\xi, \eta)$ in Γ is greater than zero. This is done as before by showing that $\rho(\xi, \eta)$ is continuous. Let (ξ_1, η_1) be a point at a distance δ from (ξ, η). We get, by an argument similar to that used above, the inequality

$$|\rho(\xi_1, \eta_1)-\rho(\xi, \eta)| < \delta,$$

and infer the continuity of $\rho(\xi, \eta)$ in Γ.

The novelty consists in the use of circles with centres (ξ, η), (ξ_1, η_1) instead of intervals with centres ξ, ξ_1.

The functions fx of a complex variable $x\,(=\xi+i\eta)$ which we shall consider will all be expressible in the form $u(\xi, \eta)+iv(\xi, \eta)$, where $u(\xi, \eta)$ and $v(\xi, \eta)$ are real functions; it follows that when fx is one-valued and continuous in a region Γ situated in the finite part of the x-plane, the continuity is uniform.

55. Uniform Convergence to a Limit. Let $f\xi$ be a continuous function of a real variable ξ in a closed interval (a, b). At each point ξ between a and b

$$\lim_{h=0} f(\xi+h)=f\xi,$$

by reason of the continuity; so that at ξ, $f(\xi+h)$ tends to the limit $f\xi$ when h tends to zero. Because the continuity is uniform it is possible to assign a positive number δ which *is independent of* ξ and such that

$$|f(\xi+h)-\lim_{h=0} f(\xi+h)| < \epsilon,$$

as soon as $|h|<\delta$. For this reason the function $f(\xi+h)$ is said to converge *uniformly* to its limit $f\xi$.

Let us consider next a more general case. Let $f(\xi, \eta;\, h, k)$ be a one-valued function of two sets of real variables ξ, η and h, k. The range of ξ, η is to be over the interior and boundary of a two-dimensional region Γ situated in the finite part of the (ξ, η)-plane. The quantities h, k, may be regarded as variable parameters which tend to given limits h_0, k_0.

Now suppose that $\lim\limits_{h=h_0,\, k=k_0} f(\xi, \eta;\, h, k)=F(\xi, \eta)$; then we know that for any assigned point (ξ, η) of Γ there can be found infinitely many numbers δ such that

$$|f(\xi, \eta;\, h, k)-F(\xi, \eta)| < \epsilon,$$

for all values of $|h-h_0|$ and $|k-k_0|$ that are less than δ. Let ρ be the upper limit of all permissible values of δ; then ρ can replace δ. This number ρ is a function $\rho(\xi, \eta)$ of ξ and η. Let the lower limit of $\rho(\xi, \eta)$ for the region Γ be Δ. Then, according as Δ is greater than or equal to zero, the function

$f(\xi, \eta; h, k)$ is said to converge uniformly or non-uniformly to its limit $F(\xi, \eta)$. Here again it is essential to notice that when $\Delta > 0$, a definite value $\delta = \Delta$ will serve for the inequality, *whatever be the position of* (ξ, η) *in* Γ.

If in the preceding work we use only one variable ξ instead of two variables ξ, η, we must replace the region Γ by an interval.

Cantor has illustrated the possibility of non-uniform convergence to a limit by means of the function $\frac{n\xi(1-\xi)}{n^2\xi^2+(1-\xi)^2}$, where ξ is real and confined to the closed interval $(0, 1)$ and n is a positive integer which tends to infinity. Take any definite value ξ in the interval; then $\underset{n=\infty}{\mathrm{L}} \frac{n\xi(1-\xi)}{n^2\xi^2+(1-\xi)^2} = 0$; but we cannot assert that there exists a positive integer μ such that when $n \geqq \mu$ we have, for *all* values ξ in the interval,

$$\frac{n\xi(1-\xi)}{n^2\xi^2+(1-\xi)^2} < \epsilon.$$

For this reason the convergence to the limit is non-uniform.

The full significance of this example can be appreciated best by tracing some curves of the family

$$\eta = \frac{n\xi(1-\xi)}{n^2\xi^2+(1-\xi)^2}$$

for values $0 \leqq \xi \leqq 1$. As the integer n grows the curves tend on the whole towards the axis of ξ; but since η has a maximum value $1/2$ when $\xi = 1/(n+1)$ each curve is a wave of height $1/2$. Let ξ take an *assigned* value ξ_0 which may be as small as we please; when n tends to infinity the parts above the interval $(\xi_0, 1)$ tend to coincidence with that interval, and ultimately the crest of the wave lies at a finite distance $1/2$ above some point of the interval $(0, \xi_0)$.

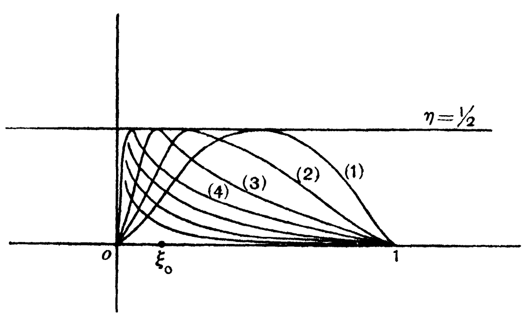


Fig. 38.

CHAPTER VII.

THE RATIONAL ALGEBRAIC FUNCTION.

56. Continuity of the Rational Integral Function. The rational algebraic function includes the rational integral function of the complex variable x, defined by

$$y = a_0 + a_1 x + a_2 x^2 + \dots\dots + a_n x^n \dots\dots\dots\dots\dots (1),$$

and the rational fractional function, or rational fraction in x, which is a quotient of two rational integral functions of x.

It is at our option whether we represent y on the x-plane as in ch. V. or use a separate y-plane as in ch. III.

An essential distinction is to be drawn between the nth power and an nth root of x. When $y = \sqrt[n]{x}$ there are, in general, n values of y for each x, i.e. y is *an n-valued function of x;* whereas the function x^n is *one-valued.* One of the first questions to be asked about any function is whether it is one-valued or not.

In the equation (1) y is evidently *one-valued* (see § 20) as regards x; we shall show presently that x is *n-valued* as regards y. Some preliminary theorems and explanations are necessary before this can be proved.

The term function is to be applied to the various expressions in x as they are introduced; the above functions will be found to be included in the general class of functions to be discussed later on under the name of analytic functions.

In this and succeeding paragraphs we shall be concerned with rational algebraic functions, including as special cases rational integral functions and mere powers; but the definitions that we shall give for continuity, limit, derivate, etc., will apply to the functions which are to be introduced successively in later chapters. The definition of continuity (§ 49) has to be modified.

The interval $(\alpha-\delta, \alpha+\delta)$ on the ξ-axis is to be replaced by the assemblage of all points x which satisfy the inequality $|x-a| \leqq \delta$, where δ is real and positive; that is, the interval, composed of all points on the real axis at a distance from a not greater than δ, is to be replaced in the plane by the circular region which contains all points x whose distances from a do not exceed δ. The definition of continuity runs now as follows:—

A function fx of a complex variable x is said to be continuous at the point $x=a$, when there exists a circular region ($|x-a| \leqq \delta$) such that at every point x of this region the difference between fx and fa is less than an arbitrarily small positive number ϵ chosen in advance. It is implied that the function is defined when $x=a$.

The following theorem serves as a basis for the proof of the continuity of the rational integral function of x:—

The rational integral function

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n.$$

which takes the value a_0 when $x=0$, can be made to assume a value as near as we please to a_0 by taking x small enough.

It is necessary to prove that $|y-a_0|$ can be made less than ϵ. We have

$$|y-a_0| = |a_1x + a_2x^2 + \dots + a_nx^n|$$
$$\leqq |x|[|a_1| + |a_2||x| + \dots + |a_n||x|^{n-1}].$$

Let μ be the greatest of the numbers $|a_1|, |a_2|, \dots, |a_n|$, then

$$|y-a_0| \leqq \mu|x|.[1 + |x| + |x|^2 + \dots + |x|^{n-1}];$$

whence, if $|x| < 1$, we have

$$|y-a_0| < \frac{\mu|x|}{1-|x|}.$$

If then we wish $|y-a_0|$ to be less than ϵ we have only to choose x so that

$$\frac{\mu\,|x|}{1-|x|}<\epsilon,$$

or

$$|x|<\epsilon/(\mu+\epsilon).$$

As such a choice of x is always possible, the theorem is proved.

Closely related with this theorem is the following:—

It is always possible to find for $|x|$ a value X such that for this and all greater values of $|x|$ we shall have

$$|a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0| > \gamma,$$

where γ is an arbitrarily great positive number which is chosen in advance.

For the rational integral function under consideration has for its absolute value

$$|x|^n\left|a_n + a_{n-1}\frac{1}{x} + a_{n-2}\frac{1}{x^2} + \ldots + a_0\frac{1}{x^n}\right|;$$

and because the absolute value of the sum of two quantities

$$a_n,\quad a_{n-1}\frac{1}{x} + a_{n-2}\frac{1}{x^2} + \ldots + a_0\frac{1}{x^n},$$

is not less than the difference of the absolute values of these quantities, it follows that the second factor in the expression for $|y|$ is not less than

$$|a_n| - \left|a_{n-1}\frac{1}{x} + a_{n-2}\frac{1}{x^2} + \ldots + a_0\frac{1}{x^n}\right|$$

But, by the preceding theorem,

$$\left|a_{n-1}\frac{1}{x} + a_{n-2}\frac{1}{x^2} + \ldots + a_0\frac{1}{x^n}\right|$$

can be made less than ϵ by making

$$\frac{1}{|x|} < \frac{\epsilon}{\mu'+\epsilon},$$

where μ' denotes the greatest of the numbers $|a_0|, |a_1|, |a_2|, \ldots, |a_{n-1}|$. Hence, when

$$|x| > (\mu'+\epsilon)/\epsilon,$$

we have

$$|y| > |x^n| \,.\, \{|a_n| - \epsilon\}.$$

To fix ideas let $\epsilon = \frac{1}{2}|a_n|$. Then

$$|y| > \tfrac{1}{2}|a_n||x|^n;$$

and we see at once that $|y| > \gamma$ for *all* values of x which make

$$|x|^n \geqq \frac{2\gamma}{|a_n|}.$$

We may take therefore for X the real positive value of $\sqrt[n]{\dfrac{2\gamma}{|a_n|}}$.

NOTE. Let x and the a's be real. Then the theorems that we have proved show respectively that the term a_0 governs the sign of the whole expression when the values of x are sufficiently small, and that the term $a_n x^n$ governs the sign when the values of x are sufficiently large.

Ex. 1. Prove that for sufficiently large values of $|x|$, the absolute value of the last term in

$$a_r x^r + a_{r+1}x^{r+1} + \ldots + a_n x^n,$$

where r is an integer which is less than n and greater than 0, is greater than the sum of the absolute values of the remaining terms.

Ex. 2. Prove that the sum and product of two functions, both continuous at a, are continuous at a. Prove the same for their ratio, when the denominator is not zero.

57. Instance and Definition of a Limit. Let us use the example of § 43 in a slightly modified form, and let x and a be complex. The function $fx \equiv (x^2 - a^2)/(x-a)$ is one-valued and is equal to $x + a$ for all values of x other than a. For $x = a$ the function is not defined at all; this does not affect the possibility of the existence of a limit when x approaches a. When $|x - a| < \delta$ we have $|(x+a) - 2a| < \epsilon$, simply by taking $\delta = \epsilon$; and therefore

$$|fx - 2a| < \epsilon$$

for all positive numbers ϵ, provided $x \neq a$ and $|x - a| < \epsilon$. We say under these conditions that fx has the *limit* $2a$; and write this,

$$\lim_{x=a} \frac{x^2 - a^2}{x - a} = 2a.$$

In general the definition of a limit is as follows. Let fx be a one-valued function of x. Denoting by ϵ, as before, an arbitrarily small positive number given in advance, then *if for every ϵ there is a corresponding positive number δ such that, when $|x - a| < \delta$ and $\neq 0$, $|fx - b| < \epsilon$, b is called the limit of fx when x tends to a.*

If for every ϵ there is a corresponding δ such that when $|x| > \delta$, $|fx - b| < \epsilon$, *then* b *is said to be the limit of* fx *when* x *tends to infinity.*

If when $|x - a| < \delta$ *and* $\neq 0$, $|fx| > 1/\epsilon$, *it is said that the limit of* fx *is infinity when* x *tends to* a.

If when $|x| > \delta$, $|fx| > 1/\epsilon$, *it is said that the limit of* fx *is infinity when* x *tends to infinity.*

Hence we see that a limit b can exist even when fa is undefined as in the case $\dfrac{x^2 - a^2}{x - a}$; further we see that in order that fx may be continuous at $x = a$, there must be one and only one limit when x tends to a, there must be a definite value fa at $x = a$, and this value must be equal to the limit.

58. The Derivate of a Function. The limit of which a particular instance has just been given is one of capital importance. A slight generalization is effected by taking x^n instead of x^2, n being a positive integer. Let then $y = x^n$, and when $x = a$ let $y = b$, so that $b = a^n$. Then

$$y - b = x^n - a^n,$$

and
$$(y - b)/(x - a) = (x^n - a^n)/(x - a).$$

Write h for $x - a$; the expression on the right-hand side becomes
$$\{(a + h)^n - a^n\}/h;$$

or, by the binomial theorem,

$$na^{n-1} + n(n-1)\,a^{n-2}\,h/2! + \ldots\ldots + nah^{n-2} + h^{n-1}.$$

By § 56 the difference between this expression and its constant term can be made as small as we please by confining $h\,(\neq 0)$ to a suitably chosen small circle about $h = 0$. Therefore by the definition of a limit

$$\lim_{x=a}\,(y - b)/(x - a) = na^{n-1}.$$

The limit thus found is called the *derivate* of x^n for the value $x = a$. Since a can take any position in the x-plane, we may replace a by x, and say that *the derivate of* x^n *at the point* x *is* nx^{n-1}.

Let us generalize and take $y = fx$, where fx is restricted

provisionally to the meaning 'a polynomial or fraction in x.' *Let fx have a definite value fa at a; the limit, if existent, of $(fx-fa)/(x-a)$ for $x=a$ is called the derivate of fx when $x=a$, or the derivate of fx at a.*

Writing x for a we have the law of the derivate for *general* values of x, excluding certain special positions. The derivate is a new function of x,—where the word function, like the symbol fx, is used only temporarily in the restricted sense of polynomial or fraction,—which is closely related to the given function; it is denoted sometimes by an accent $f'x$, sometimes by the capital initial letter D of 'derivate, Dfx. A third notation is dfx/dx; the origin of this notation, and of its accompanying name the differential quotient, may be worth explaining.

The definition of the derivate is in words "the limit of the ratio $\dfrac{\text{change of function}}{\text{change of variable}}$, when the change of the variable tends to zero." Call this limit,—supposed to exist,—Dfx. We define the differential of a function for an assigned x by the equation

$$dfx = Dfx \times (\text{an arbitrary change of } x).$$

If the function be merely x itself, $Dfx = 1$ and the differential of x, namely dx, is simply the arbitrary change of x. The equation may be rewritten therefore as follows.

$$dfx = Dfx \times dx.$$

Thus the ratio or quotient of the differentials dfx, dx is the same as the derivate Dfx.

The formal rules for constructing the derivate of fx are the same whether we use complex or real values. Thus, when the definition of a function applies to complex values of x as much as to real, the derivate will have the same form in both cases. The existence of a derivate which is not zero ensures isogonality (§ 26).

Taylor's theorem for the rational integral function, namely that

$$fx = fa + f'a\,(x-a) + f''a\,\frac{(x-a)^2}{2!} + \dots + f^{(n)}a\,\frac{(x-a)^n}{n!},$$

is proved precisely as for real variables. The theorem enables

us to arrange a series of powers of x in a series of powers of $x-a$, in other words to change the origin.

59. Fundamental Theorem of Algebra. The fundamental theorem to which we refer is that *every equation*

$$fx \equiv a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1} + a_n x^n = 0,$$

in which the a's are complex (real values of course included) and $a_n \neq 0$, *has a root.*

With this proposition stands or falls the associated theorem that *every algebraic equation has n roots.*

It is obvious that if the first theorem be not true, the second is not true; let us assume then that the first theorem is true and prove that the second follows from it.

Let $fx = 0$ be satisfied by $x = x_1$. Then

$$fx = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n,$$

$$0 = a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_n x_1^n;$$

hence $\quad fx = a_1(x - x_1) + a_2(x^2 - x_1^2) + \ldots\ldots + a_n(x^n - x_1^n).$

As every term on the right-hand side is divisible by $x - x_1$, fx must be divisible by $x - x_1$. Let us write then

$$fx = (x - x_1)(b_0 + b_1 x + b_2 x^2 + \ldots + b_{n-1} x^{n-1}) = (x - x_1) f_1 x,$$

where the second factor is necessarily of degree $n - 1$. By hypothesis, $f_1 x = 0$ is satisfied by some value x_2; a repetition of the preceding argument shows that

$$f_1 x = (x - x_2) f_2 x,$$

where $f_2 x$ is a rational integral function of degree $n - 2$. By continuing this process we arrive finally at the formula

$$fx = a_n (x - x_1)(x - x_2) \ldots (x - x_n).$$

This formula shows at once that fx vanishes for $x = x_1, x_2, \ldots, x_n$ and for no other values of x.

In the next article we shall give a rigorous analytic proof that every algebraic equation has at least one root; in the present one we adduce some geometric considerations which throw light on the analysis.

When x traverses the whole of the x-plane, y must traverse

the whole or part of the y-plane; in the former case one of the values of y is 0, in the latter there are certain regions of the y-plane which are never reached and one of these regions may contain $y=0$.

When x is at any point a, such that $f'a \neq 0$, let y be at b. If b is zero, the equation has a root; suppose then that $b \neq 0$. Then y can be brought nearer to the origin than b is. For by the property of isogonality, angles at b are equal to angles at a; hence by a proper choice of the direction at a we can make the direction at b what we please.

The condition $f'a \neq 0$ ensures isogonality. The points a excluded by this condition are at most $n-1$; these points, being finite in number, can be avoided. Thus we can bring y nearer and nearer to the origin; but this does not show that y reaches the origin.

60. Proof of the fundamental Theorem. We have proved that there exists a circle, with origin as centre, which is such that

$$|a_0 + a_1x + a_2x^2 + \dots + a_nx^n| > \gamma$$

for all points on and exterior to this circle, γ being an arbitrary positive number assigned in advance. All the roots of the algebraic equation, in case they exist, must lie within this circle.

Let a be a finite value of x and let $b = fa$. Suppose that $b \neq 0$.

I. We propose to prove in the first place that there is a value $a+h$ in the x-plane which gives a point y whose distance from the origin is less than that of b; that is, which makes

$$|f(a+h)| < |fa|, \text{ or } \left|\frac{f(a+h)}{fa}\right| < 1.$$

To take the most general case assume that in Taylor's expansion the coefficients $f'a, f''a, \dots, f^{r-1}a$ vanish, while $f^ra \neq 0$. Then

$$\frac{f(a+h)}{fa} = 1 + \frac{h^r}{r!}\frac{f^ra}{fa} + \frac{h^{r+1}}{r+1!}\frac{f^{r+1}a}{fa} + \dots + \frac{h^n}{n!}\frac{f^na}{fa}.$$

Let $\quad h = \rho(\cos\theta + i\sin\theta), \quad \dfrac{1}{s!}\dfrac{f^sa}{fa} = \rho_s(\cos\theta_s + i\sin\theta_s),$

where $s = r, r+1, \ldots, n$; then

$$\frac{f(a+h)}{fa} = 1 + \rho^r\rho_r[\cos(r\theta+\theta_r) + i\sin(r\theta+\theta_r)]$$

$$+\rho^{r+1}\rho_{r+1}[\cos(\overline{r+1}\,\theta+\theta_{r+1}) + i\sin(\overline{r+1}\,\theta+\theta_{r+1})] + \ldots$$
$$+\rho^n\rho_n[\cos(n\theta+\theta_n) + i\sin(n\theta+\theta_n)].$$

Now let θ be so chosen as to make

$$r\theta + \theta_r = \pi;$$

then $$\frac{f(a+h)}{fa} = 1 - \rho^r\rho_r + \text{terms in } \rho^{r+1}, \rho^{r+2}, \ldots, \rho^n,$$

whose absolute values are $\rho^{r+1}\rho_{r+1}, \rho^{r+2}\rho_{r+2}, \ldots, \rho^n\rho_n$; and let ρ be sufficiently small to secure the inequality $\rho^r\rho_r < 1$.

The absolute value of the expression on the right-hand side cannot exceed

$$1 - \rho^r\rho_r + \rho^{r+1}\rho_{r+1} + \ldots + \rho^n\rho_n;$$

that is,

$$\left|\frac{f(a+h)}{fa}\right| \leqq 1 - \rho^r\rho_r\left[1 - \rho\frac{\rho_{r+1}}{\rho_r} - \rho^2\frac{\rho_{r+2}}{\rho_r} - \ldots - \rho^{n-r}\frac{\rho_n}{\rho_r}\right].$$

By the theorem of § 56, the expression

$$1 - \rho\frac{\rho_{r+1}}{\rho_r} - \rho^2\frac{\rho_{r+2}}{\rho_r} - \ldots - \rho^{n-r}\frac{\rho_n}{\rho_r}$$

can be made to differ from unity by as little as we please, provided we choose ρ sufficiently small. It is therefore possible to find a value of h such that

$$\rho^r\rho_r\left[1 - \rho\frac{\rho_{r+1}}{\rho_r} - \ldots\right]$$

is positive and less than unity. For this value we have

$$|f(a+h)| < |fa|.$$

The proposition that we have established can be stated as follows:—

Given that $|fx|$ does not vanish for an assigned value $x = a$, it is always possible to alter x so as to diminish $|fx|$.

II. Secondly let us consider the lower limit γ' of the values of $|fx|$. Since $|fx|$ is a continuous function of the real variables ξ, η, this lower limit is attained and is therefore a minimum

value of $|fx|$ (§ 52). If possible let this minimum value γ' be different from zero. Since there is a value x for which $|fx|$ is γ' exactly, it follows from what we have shown above that it must be possible to find a value of $|fx|$ which is less than γ', contrary to the supposition that γ' is the lower limit of the values $|fx|$. Hence γ' must be 0, and the value 0 is attained. This proves that there exists a value of x which makes fx vanish.

NOTE. The theorem that the values of $|fx|$ can be made smaller and smaller by successive alterations of x is not, by itself, sufficient to establish the existence of an x which makes $fx = 0$. The initial value $|fa|$ might for example be 3 and the diminishing values produced by the successive alterations might be $2\frac{1}{2}$, $2\frac{1}{3}$, $2\frac{1}{4}$, $2\frac{1}{5}$,..., a sequence whose lower limit is 2. As soon, however, as we are certain, from our knowledge of the properties of continuous functions of two real variables, that the lower limit of this sequence is attained by $|fx|$ for some value of x, we see that still smaller values can be attained, and that 2 is not the lower limit of the complete system of values of $|fx|$. Thus the lower limit of this complete system is 0, and drawing again on the above-mentioned property of the continuous function $|fx|$, we know that the value 0 is attained.

Here is a concrete example of the necessity of examining whether a variable quantity does or does not attain its lower limit.

Since every algebraic equation of degree n in x is satisfied by n values of x it follows immediately that the equation

$$fx - y = 0$$

gives n values of x for each value of y; this justifies the statement of § 56.

61. The Rational Algebraic Function of x. Let

$$y = fx \equiv (a_0 + a_1x + a_2x^2 + \dots + a_nx^n)/(b_0 + b_1x + b_2x^2 + \dots + b_mx^m);$$

it is evident that y is in general a one-valued function of x. To make this statement universally true we suppose the fraction in its lowest terms; further we must *assign* one value to y when $x = \infty$, and regard $y = \infty$ as a value as in § 20. "The value of y when $x = \infty$" is in itself meaningless, because ∞ is not a definite number, but $\lim_{x=\infty} y$ often exists and then we define $\text{value}_{x=\infty} y$ to mean $\lim_{x=\infty} y$. Thus in the present case y has one value when $x = \infty$;

for when $|x|$ is sufficiently large $a_n x^n$, $b_m x^m$ become the all-important terms of the numerator and denominator and their ratio tends to ∞, a_n/b_m, or 0, according as $n >, =, < m$.

When $n > m$ we can divide the denominator into the numerator, getting a quotient $c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-m} x^{n-m}$ and a remainder $d_0 + d_1 x + d_2 x^2 + \ldots + d_{m-1} x^{m-1}$. Hence

$$y = c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-m} x^{n-m} + \frac{d_0 + d_1 x + d_2 x^2 + \ldots + d_{m-1} x^{m-1}}{b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m}.$$

Partial Fractions. If $x_1, x_2, \ldots, x_k$ be the roots of the equation formed by equating the denominator to zero, these roots occurring to the orders of multiplicity $m_1, m_2, \ldots, m_k$, so that $m_1 + m_2 + \ldots + m_k = m$; then, as we know, the fraction can be resolved into partial fractions and y takes the form

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-m} x^{n-m}$$
$$+ \sum_{r=1}^{k} \left[\frac{A_{r1}}{x - x_r} + \frac{A_{r2}}{(x - x_r)^2} + \ldots + \frac{A_{rm_r}}{(x - x_r)^{m_r}} \right].$$

When $n = m$, the part external to the sign of summation dwindles to a_n/b_m, and when $n < m$ it disappears altogether.

The Infinities of the Rational Fraction. By an *infinity* we mean a value x which makes the function infinite; and by a *zero* we mean a value x which makes the function vanish. When $x = x_1, x_2, \ldots, x_k$ the rational fraction is infinite. It is evident that y will, when multiplied by $(x - x_r)^{m_r}$, have a finite limit at $x = x_r$, but the limit infinity when multiplied by any lower power of $x - x_r$. This being so we define *the order of infinity* at a point x' as follows :—

The rational fraction fx has an infinity of order s, where s is a positive integer, at a point x' when the limit of $fx.(x - x')^s$ is finite at $x = x'$, while the limit of $fx.(x - x')^{s-1}$ is infinite.

This definition applies also to other one-valued functions to be introduced later.

We have assumed tacitly that x' is finite. Suppose however that $x' = \infty$, and that y is infinite for this value. Then $n > m$; and the part

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_{n-m} x^{n-m}$$

exists and its most important term is $c_{n-m}x^{n-m}$ when $|x|$ is large. We say that $n-m$ is the *order of infinity of y when $x=\infty$*. Let us frame the definition in such a way that it may be applicable also to other one-valued functions than the rational. If we divide by x^{n-m} the function ceases to be infinite when $x=\infty$; it takes in fact the value c_{n-m}. But the function, when divided by any lower power of x than the $(n-m)$th, is infinite when $x=\infty$. We say that *the function has an infinity, at $x=\infty$, of order s, where s is a positive integer, when the function fx/x^s is finite for $x=\infty$, while fx/x^{s-1} is infinite for the same value.*

The Zeros of the Rational Fraction. Let

$$fx=\frac{a_n(x-x_1')^{n_1}(x-x_2')^{n_2}\dots(x-x_h')^{n_h}}{b_m(x-x_1)^{m_1}(x-x_2)^{m_2}\dots(x-x_k)^{m_k}},$$

where $n_1+n_2+\dots+n_h=n$ and $m_1+m_2+\dots+m_k=m$; we say that the zeros at x_1', x_2', ..., x_h' are of orders n_1, n_2, ..., n_h respectively, just as the infinities were said to be of orders m_1, m_2, ..., m_k. When $m>n$ the function has a zero at $x=\infty$; this zero is said to be of the order $m-n$, just as when $m<n$ the function has an infinity, at $x=\infty$, of the order $n-m$.

By addition we see at once that *the sum of the orders of the zeros ($x=\infty$ included, if a zero) is equal to the sum of the orders of the infinities ($x=\infty$ included, if an infinity).*

For example, the sum of the orders of the zeros of

$$\frac{x^4+x^2+1}{x^6+x+1}$$

is 6; the zeros in the finite part of the plane contribute 4 and $x=\infty$ contributes 2. The sum of the orders of the infinities of the fraction is 6, for the denominator vanishes at six points.

CHAPTER VIII.

CONVERGENCE OF INFINITE SERIES.

62. Dennition of an Infinite Series. When the four elementary operations are applied a finite number of times to rational functions of x, the resulting functions are of the same kind. To get anything new we must use the notion of a *limit.*

In the present chapter we discuss the questions raised by the unlimited repetition of the process of addition, or what is called the theory of infinite series. The reader has of course examined some special infinite series, and made use of approximations to infinite series; for example in calculating logarithms, or the number π. But it seems best to take up the general theory from the beginning, without discussing details which will not concern us later.

Ex. As we have especially in view operations with complex numbers the reader is advised to draw the strokes which represent the sums of the first few terms in the following simple cases:

$$1+i/2!+i^2/3!+i^3/4!+\ldots,$$
$$1+i/2\ +i^2/2^2\ +i^3/2^3+\ldots,$$
$$1+i/2\ +i^2/3\ \ +i^3/4\ \ +\ldots,$$
$$1+i\ \ \ +i^2\ \ \ \ +i^3\ \ \ \ +\ldots;$$

examining if and how the sums tend to a limit, and considering how to calculate the limit when it exists.

We have first to assign a meaning to the word *series.* When we consider a succession of operations we ought either to attend to their order,—what is to be done first, what second, and so on,—

or to show that this is needless. We may contrast as simple instances $1+2+3$ or $3 \times 4 \times 5$, where elementary theorems tell us that the order does not matter, with $2/2/2$ which might mean 2 or $\frac{1}{2}$; or 3^{3^3} which might mean 3^{27} or 27^3.

When the terms of a sum are to be added in the order in which they are written, the operation is called a series.

This definition requires the addition to be performed in a particular order. Thus when we speak of the series $1+2+3+4$ we understand that we are to add 2 to 1, 3 to this partial sum, and 4 to this partial sum. We thus form a succession of numbers, the last of which is the number sought. This succession of numbers might itself be called a series; and so might a product be called a series when we attend to the order of the operations, but the word series is used in mathematics with especial reference to addition. Thus for the succession of numbers 1, 3, 6, 10 of this example we retain the word sequence, and we observe that the series is the process or series of simple operations by which we build the sequence from the terms.

We are so familiar with the association and commutation of terms in adding an assigned number of terms that in this case the explanation may appear trivial. The importance of what we have said lies in its bearing on the case of an unlimited number of terms; the operation is then not the building of a number, but the building of an infinite sequence. Such an operation is called an *infinite series.*

Thus if the terms in the first, second, ..., nth place be $a_1, a_2, \ldots, a_n$, and we write

$$s_n = a_1 + a_2 + \ldots + a_n,$$

we build the sequence of sums

$$s_1, s_2, \ldots, s_n, \ldots,$$

and the act (whether of thought, writing, or speech) by which we pass from the sequence of terms to the sequence of sums,—this is the infinite series.

63. Convergence. The sequence $s_1, s_2, \ldots, s_n, \ldots$ may tend to a finite limit s. When and only when this is the case

we say that *the series is convergent.* In this case the series leads to a number, namely s. When there is not a finite limit the series is *divergent.* Mathematical notation is concerned especially with what results from operations. When we write $a+b$, the operation is the replacing of two numbers by a single one; the notation, so soon as we are used to it, suggests that single number. A notation introduced to signify an operation ends usually by signifying the outcome.

So in this case we first agree to denote by

$$a_1 + a_2 + \dots + a_n + \dots,$$

by $\sum\limits_{n=1}^{\infty} a_n$, or by Σa_n, the series or operation itself, and then when the limit s exists we agree to use the same notation for s itself. Lastly, as an almost unavoidable result of using the same notation for both the limit and the operation, the limit of a convergent series is frequently spoken of as the series itself. But it must not be overlooked that when there is no limit we must recur to the definition of the series.

NOTE. In most English text-books a series is said to be divergent only when it tends to ∞. Thus $1-1+2-2+3-3+\dots$ is (in this view) neither convergent nor divergent. The definition adopted here has the support of Cayley (*Encyc. Brit.*, Art. Series) and Stokes (*Camb. Phil. Trans.* vol. viii. p. 535), among English authorities, and is prevalent elsewhere.

We use L to signify the limit when n tends to ∞, if this limit is known to exist.

We mean by a real series one with real terms alone; by a complex series one with complex terms.

A necessary and sufficient condition for the convergence of Σa_n *is that corresponding to* **every** *positive number* ϵ *given in advance there shall exist a positive integer* μ *such that the absolute value of* $a_{n+1} + a_{n+2} + \dots + a_{n+p}$ *shall be less than* ϵ *for every integer* n *equal to or greater than* μ, *and for every positive integer* p.

That is, $|s_{n+p} - s_n| < \epsilon$ when $n \geqq \mu$. See § 47.

The one integer μ is to serve for every selection of $n\,(\geqq \mu)$ and of p. If p were assigned, instead of being free, the condition

of convergence, though necessary, would cease to be sufficient. For instance in the series $1 + 1/2 + 1/3 + 1/4 + \ldots$, the sum of an assigned number of terms following the nth can be made as small as we please by increasing n; whereas the sum of n terms following the nth is at once seen to be greater than $1/2$.

When in calculating the limit of a convergent real series we decide on a degree of approximation,—say we want to be within $1/10^6$ of the limit,—the least corresponding number μ measures the rapidity of the convergence.

The expression $s - s_n$ is called the *remainder* or *residue* of a convergent series. It is of course itself a limit; it will be denoted by r_n. Since

$$s = s_n + r_n,$$

and $Ls_n = s$, it follows that $Lr_n = 0$.

64. Simple tests of Convergence for Series with terms all positive. Whether a series is convergent or not can be decided in some simple but important cases by a comparison term by term with a standard series. An especially convenient standard series is the geometric series $\sum_{n=0}^{\infty} x^n$. This is convergent when x is any number, real or complex, such that $|x| < 1$. For since

$$s_n = \frac{x - x^{n+1}}{1 - x},$$

we have

$$\left| s_n - \frac{x}{1-x} \right| = \frac{|x|^{n+1}}{|1-x|} \leqq \frac{|x|^{n+1}}{1 - |x|},$$

or if $|x| = \dfrac{1}{1+\alpha}$, where $\alpha > 0$,

$$\left| s_n - \frac{x}{1-x} \right| \leqq \frac{1}{\alpha(1+\alpha)^n} < \frac{1}{\alpha(1+n\alpha)};$$

and $\dfrac{1}{\alpha(1+n\alpha)} < \epsilon$ when $n > \dfrac{1-\alpha\epsilon}{\alpha^2\epsilon}$.

Here then μ is any integer $> \dfrac{1-\alpha\epsilon}{\alpha^2\epsilon}$, and there is the limit

$$s = \frac{x}{1-x}.$$

(1) *Let Σa_n and $\Sigma a'_n$ be series with terms all positive and let Σa_n be convergent with the limit s. If $a'_n \leqq a_n$, from an assigned value of n onwards, then also $\Sigma a'_n$ is convergent.*

For if s'_n denote the sum of n terms of the accented series we have

$$\begin{aligned} s'_n - s'_m &\leqq s_n - s_m \\ &< s - s_m; \end{aligned}$$

or
$$s'_n < s - s_m + s'_m;$$

we have then the numbers s'_n which increase with a but do not tend to ∞. There is then (§ 45) a limit s'.

Ex. So if Σa_n is divergent and from a given number m onwards $a'_n \geqq a_n$, $\Sigma a'_n$ is divergent.

(2) *If Σa_n is convergent and if when $n \geqq m$*

$$\frac{a'_{n+1}}{a'_n} \leqq \frac{a_{n+1}}{a_n},$$

then $\Sigma a'_n$ is convergent.

For if
$$\frac{a'_{m+1}}{a'_m} \leqq \frac{a_{m+1}}{a_m},$$

$$\frac{a'_{m+2}}{a'_{m+1}} \leqq \frac{a_{m+2}}{a_{m+1}},$$

and so on, then by multiplication,

$$\frac{a'_n}{a'_m} \leqq \frac{a_n}{a_m}, \text{ when } n \geqq m.$$

Since $\sum_{n=1}^{\infty} a_n$ is convergent, so is $\sum_{n=m}^{\infty} a_n$ and so is $\sum_{n=m}^{\infty} a_n/a_m$.

Hence, by the preceding theorem, $\sum_{n=m}^{\infty} a'_n/a'_m$ is convergent; and therefore so are $\sum_{n=m}^{\infty} a'_n$ and $\sum_{n=1}^{\infty} a'_n$.

Ex. If Σa_n is divergent and, from an assigned value of n onwards, $\frac{a'_{n+1}}{a'_n} \geqq \frac{a_{n+1}}{a_n}$, then $\Sigma a'_n$ is divergent.

(3) *When $\mathrm{L}\; a_{n+1}/a_n$ exists and is less than 1, the series is convergent.*

Let the limit be $1-\alpha$ where α lies between 0 and 1; then to every ϵ there is a μ such that when $n \geqq \mu$,

$$1-\alpha-\epsilon<\frac{a_{n+1}}{a_n}<1-\alpha+\epsilon.$$

Choose a definite ϵ, $<\alpha$, so that $1-\alpha+\epsilon<1$. Let the corresponding μ be m; then comparing a_n with a geometric series whose constant ratio is $1-\alpha+\epsilon$ we see that the conditions of the preceding theorem are satisfied.

Ex. When $\mathrm{L}\,\frac{a_{n+1}}{a_n}$ exists and is greater than 1, the series is divergent. If the limit is 1 or if there is no limit, other tests must be used. See § 76.

(4) *When* $\mathrm{L}\,a_n^{1/n}$ *exists and is less than* 1, *the series is convergent.*

For proceeding as in (3) we have

$$a_n^{1/n}<1-\alpha+\epsilon<1,\ n \geqq m.$$

Since $a_n<(1-\alpha+\epsilon)^n$ and since the geometric series whose nth term is $(1-\alpha+\epsilon)^n$ is convergent, the conditions of (1) are satisfied.

65. Association of the Terms of a Series. By the sum of two series Σa_n and $\Sigma a'_n$, real or complex, we understand $\Sigma(a_n+a'_n)$. It is clear that, if $|s_{n+p}-s_n|<\epsilon$ and $|s'_{n+p}-s'_n|<\epsilon$, then $|s_{n+p}+s'_{n+p}-s_n-s'_n|<2\epsilon$ (§ 14). Hence the limit of $\Sigma(a_n+a'_n)$ is the sum of the limits of Σa_n, $\Sigma a'_n$.

The theorem that $|\sum_1^n a_n| \leqq \sum_1^n |a_n|$ is of great use in handling series. Observe that if $\Sigma|a_n|$ has a limit S all that the theorem tells us is that if also $|\Sigma a_n|$ has a limit $|s|$, then $|s| \leqq S$. But applying the theorem to p terms after the nth we have

$$|s_{n+p}-s_n| \leqq |S_{n+p}-S_n|,$$

where S_n is the sum of the first n terms of $\Sigma|a_n|$; whence we obtain the much more important information that if $\Sigma|a_n|$ has a limit S then Σa_n must have a limit s.

Suppose that the consecutive terms of a convergent series Σa_n are associated in sets without displacements of the a's and

that these sets are replaced by their sums $b_1, b_2, b_3, \ldots$, where b_m denotes the sum of the terms in the mth set, then $\Sigma b_m = \Sigma a_n$. For let

$$s_n = a_1 + a_2 + a_3 + \ldots + a_n, \quad t_m = b_1 + b_2 + b_3 + \ldots + b_m,$$

and let a_n be the last a in the mth set; then n tends to infinity with m and the equation $t_m = s_n$ carries with it

$$\mathrm{L}\, t_m = \mathrm{L}\, s_n.$$

It is then permissible to introduce parentheses in an infinite series. It is not always permissible to remove them; for example in the series

$$(1-1)+(1-1)+(1-1)+\ldots\ldots,$$

each term is 0 and the limit is 0, but the series

$$1 - 1 + 1 - 1 + \ldots$$

oscillates*; if we agree to take always an even number of terms the limit is 0; if an odd number, the limit is 1.

As an example of the theorem we prove that $\Sigma 1/n^2$ is convergent. The series may be written

$$1 + (1/2^2 + 1/3^2) + (1/4^2 + 1/5^2 + 1/6^2 + 1/7^2) + \ldots$$

and the terms so written are less, after the first, than those of

$$1 + (1/2^2 + 1/2^2) + (1/4^2 + 1/4^2 + 1/4^2 + 1/4^2) + \ldots$$

i.e. are less than the terms of $1 + 1/2 + 1/4 + \ldots$.

When we have associated the terms of a series in sets we can rearrange the terms in each set; we thus get a new convergent series with the same limit. But this is a special rearrangement of a series. As an instance of rearrangement which in fact gives a new series with a different limit, consider the logarithmic series $\Sigma(-1)^n/n$, or

$$1 - 1/2 + 1/3 - 1/4 + \ldots \quad \ldots\ldots\ldots\ldots\ldots\ldots(1).$$

First we prove that the series is convergent. For s_n is positive, as appears from writing it in the form

$$(1 - 1/2) + (1/3 - 1/4) + \ldots,$$

* Most English text-books regard oscillating series as not divergent (Note, § 63).

where the last term is positive whether n be odd or even; and $s_n < 1$, as appears from writing it in the form

$$1-(1/2-1/3)-(1/4-1/5)-\ldots,$$

where each term after the first is negative whether n be odd or even. Hence s_{2n} is a number which increases with n and does not become ∞; therefore it has (§ 45) a limit s; and s_{2n+1} is a number which diminishes with n and does not become $-\infty$, therefore it has a limit s'. And we have to show that $s = s'$.

Now $\qquad s_{2n+1} - s_{2n} = 1/(2n+1)$.

so that $\qquad \mathrm{L}\,(s_{2n+1} - s_{2n}) = 0$,

giving $\qquad \mathrm{L}\,s_{2n+1} = \mathrm{L}\,s_{2n}$, and $s' = s$.

The series is therefore convergent.

Now rearrange the above series, taking in turn two positive terms and one negative term, but keeping the positive, and negative, terms in the same decreasing order. We have the new series

$$1 + 1/3 - 1/2 + 1/5 + 1/7 - 1/4 + \ldots \quad \ldots\ldots(2),$$

and we do not know whether this has a limit or whether, if it has, the limit is the same, namely Log 2. For it must be observed that the rearrangement is not associative; the first n places of the new series do not contain the terms which filled the first n places of the old series, when $n > 3$.

Take the terms by threes, that is, consider the series

$$(1+1/3-1/2)+(1/5+1/7-1/4)+\ldots \quad \ldots\ldots(3).$$

We prove first that this third series has a limit.

Let α_n denote $\qquad 1 + 1/2 + 1/3 + \ldots + 1/n$.

Then $\alpha_n/2$ is $\qquad 1/2 + 1/4 + \ldots + 1/2n$,

and $\alpha_{2n} - \alpha_n/2$ is $\quad 1 + 1/3 + 1/5 + \ldots + 1/(2n-1)$;

so that if s_{2n} is, as above,

$$1 - 1/2 + 1/3 - \ldots - 1/2n,$$

$$s_{2n} = \alpha_{2n} - \alpha_n.$$

The sum of $3n$ terms of (2), that is of n terms of (3), is

$$\begin{aligned} s'_{3n} &= 1 + 1/3 + \ldots - 1/(4n-1) - (1/2 + 1/4 + \ldots + 1/2n) \\ &= \alpha_{4n} - \alpha_{2n}/2 - \alpha_n/2 \\ &= s_{4n} + s_{2n}/2. \end{aligned}$$

Hence
$$\mathrm{L}s'_{3n} = \mathrm{L}s_{4n} + \mathrm{L}s_{2n}/2$$
$$= \mathrm{Log}\, 2 + \tfrac{1}{2}\,\mathrm{Log}\, 2$$
$$= \tfrac{3}{2}\,\mathrm{Log}\, 2.$$

Thus (3) is convergent, but has not the same sum as (1). We have now to see whether we may remove the parentheses in (3) so as to get (2); that is whether (2) has the same sum as (3). The sum of $3n+1$ or of $3n-1$ terms of (2) differs from the sum of $3n$ terms by $1/(4n-1)$ or by $1/2n$; that is by a number which has itself the limit zero. Thus in the series (2) the limit is the same when we take three terms at a time, whether we begin our sets of 3 with the first, second, or third term; but this amounts to taking the terms one by one.

Ex. 1. Prove, by means of ϵ and μ, that when $\mathrm{L}a_n = 0$ and the series
$$(a_1+a_2)+(a_3+a_4)+\ldots$$
is convergent, so is the series Σa_n.

Ex. 2. Prove that if we take first p positive terms of the series for Log 2, in their natural order, and then q negative terms and repeat this process indefinitely, the sum is $\mathrm{Log}\, 2 + \tfrac{1}{2}\,\mathrm{Log}\, p/q$. The easiest way to prove this is to assume Euler's theorem that $1 + 1/2 + 1/3 + \ldots + 1/n - \mathrm{Log}\, n$ has a finite limit when n is ∞. See § 109.

66. Absolutely Convergent Series. Let Σa_n be any convergent series; and let the terms be rearranged. When the first series is finite, the two series have the same sum; when the two series are infinite (if one is, so is the other) it is by no means necessary that they have the same limit; and it is of cardinal importance to know when this is the case.

First let Σa_n be a convergent series with positive terms.

Let the new series be $\Sigma a'_n$; that is, the terms
$$a_1,\ a_2,\ a_3,\ \ldots,\ a_n,\ \ldots$$
are rearranged in the order
$$a'_1,\ a'_2,\ a'_3,\ \ldots,\ a'_n,\ \ldots;$$
the term in every assigned place in the old series has a definite place in the new; and conversely. There is to be no omission and no repetition; but for the same term there is to be a one-to-one correspondence of the old place and the new.

Take now n terms of the old series; these will be found in the new series, say they are in the first m terms $a'_1, a'_2, \ldots, a'_m$, where $m \geqq n$. And all these m terms will be found in the old series, say in the first $n+p$ terms, where $n+p \geqq m$. Hence if s_n and s'_n denote the sums of n terms of the two series,

$$s_n \leqq s'_m, \text{ and } s'_m \leqq s_{n+p}.$$

Since then $\mathrm{L}\, s_n = \mathrm{L}\, s_{n+p}$, $= s$ say, $\mathrm{L}\, s'_m$ is neither greater nor less than s, that is, $\mathrm{L}\, s'_m = s$.

Next let Σa_n be any convergent series, and let A_n be the absolute value of a_n. Let the terms be rearranged as before into $a'_1 + a'_2 + a'_3 + \ldots$; then as before the first n terms of the old series are contained in the first m terms of the new, and these in the first $n+p$ of the old, where

$$n \leqq m \leqq n+p.$$

We can no longer assert that $s_n \leqq s'_m$, but we can assert that the terms which are in s'_m, but not in s_n, are all among the terms $a_{n+1}, a_{n+2}, \ldots, a_{n+p}$. That is $s'_m - s_n$ is made up of some or all of the terms in $s_{n+p} - s_n$; and therefore $|s'_m - s_n| \leqq$ the sum of some or all of the numbers $A_{n+1}, A_{n+2}, \ldots, A_{n+p}$, and *a fortiori* $\leqq$ the sum of all these numbers. If then $A_{n+1} + A_{n+2} + \ldots + A_{n+p}$ can be made as small as we please by suitably selecting n, we can assert that

$$|s'_m - s_n| < \epsilon,$$

and therefore that $\qquad \mathrm{L}\, s'_m = \mathrm{L}\, s_n.$

We are thus led to a classification of series. For any convergent series

$$|a_{n+1} + a_{n+2} + \ldots + a_{n+p}| < \epsilon;$$

but if further the absolute values of the terms form a convergent series, then

$$A_{n+1} + A_{n+2} + \ldots + A_{n+p} < \epsilon.$$

Such series are said to be *absolutely convergent*; and what is proved is that *the terms of an absolutely convergent series can be rearranged without altering the limit.* Thus an absolutely convergent series, at first sight, might as well be called an infinite sum as it has the essential commutative characteristic of the common or finite sum. But a sufficient reason for retaining the

word series lies in the fact that in the most useful series the nth term is a simple expression in n, so that there is a natural order; and again when the terms contain x the convergence depends on x and is by no means always absolute when it exists.

67. Conditionally Convergent Series. We have now to prove a theorem completing the one just proved: namely that when a series is convergent, but not absolutely, it cannot be rearranged at pleasure without altering the sum.

First, a real convergent series, which is not absolutely convergent, must evidently contain an unlimited number of both positive and negative terms. And the series formed by the terms with the same sign must be divergent; otherwise the original series would be the difference of absolutely convergent series and would be absolutely convergent.

Now such a series can by rearrangement be made to have any sum S,—a point first noticed by Riemann. For suppose S positive; and take first positive terms, in their order, stopping when the sum is first greater than S; then add on negative terms, in their order, stopping when the sum is first less than S, and so on. Thus the sum will oscillate about S, differing from it at each stage by not more than the last term taken. Since the terms of the series tend to zero, the sum taken in this way has the limit S. The process is the same when S is negative.

Second, in the case of a series with complex terms, such a series, whose nth term is $\alpha_n + i\beta_n$, is at once reduced to $\Sigma\alpha_n + i\Sigma\beta_n$. It is convergent only when $\Sigma\alpha_n$ and $\Sigma\beta_n$ are themselves convergent. If it is convergent, but not absolutely, then $\Sigma\,|\,\alpha_n + i\beta_n\,|$ is divergent. But

$$|\,\alpha_n + i\beta_n\,| \leqq |\,\alpha_n\,| + |\,\beta_n\,|.$$

Hence $\qquad\qquad \Sigma\,(|\,\alpha_n\,| + |\,\beta_n\,|)$ is divergent.

Thus one at least of the series $\Sigma\alpha_n$, $\Sigma\beta_n$ is not absolutely convergent. And hence either the real part, or the imaginary part, of the sum of the proposed series can be made to take an arbitrary value by a suitable rearrangement in Riemann's manner. In special cases both parts can take arbitrary values.

Ex. Prove that the sum of

$$i - i^2/2 + i^3/3 - i^4/4 + \dots$$

can be made, by a suitable rearrangement, to represent any point of the plane.

A series which is convergent but can have different sums when rearranged, is said to be *conditionally convergent*; and what we have proved is that *all convergent series are either absolutely or conditionally convergent.* For we proved in the last article that when convergence is absolute it is not conditional, and in this article that when convergence is not absolute it is conditional.

68. Conversion of a Single Series into a Double Series. There remains a very important question with regard to series; namely if we break a series up into a number of *infinite* series, when can we be sure that the sum of the sums of these series is the sum of the original series? And especially what happens when we break up the series into infinitely many infinite series? These questions are not answered by what was said on rearrangements in the preceding articles. For to consider the simplest case, when we first take the limit of the terms $a_1 + a_3 \dots + a_{2n-1} + \dots$, and then the limit of $a_2 + a_4 + \dots + a_{2n} + \dots$, this is no rearrangement in the sense of § 66. For in rearranging the given series we should only have the terms $a_1, a_3, \dots$ and the terms with even suffixes would be entirely omitted. This special case is however merely that of the sum of two series; but nothing that has been said meets the following case.

Take first a_1, then the sum of all a's with even suffixes, then the sum of all remaining terms whose suffixes are multiples of three, then the sum of all remaining terms whose suffixes are multiples of five, and so on for the successive primes, on the principle of the Sieve of Eratosthenes. We have then an instance of the breaking up of a series into infinitely many infinite series; and this we now discuss in the case when the original series is absolutely convergent.

Let $a_1, a_2, a_3, \dots, a_n, \dots$ be an infinite sequence of numbers. There are many ways of arranging these numbers in the form of

a rectangular array, or *a table of double entry* as it is often called. For example using the method of the sieve we can write down the following array :—

$$\begin{array}{llllll}
a_1 & a_2 & a_4 & a_6 & a_8 & a_{10}\ldots\ldots \\
a_3 & a_9 & a_{15} & a_{21} & a_{27} & a_{33}\ldots\ldots \\
a_5 & a_{25} & a_{35} & a_{55} & a_{65} & a_{85}\ldots\ldots \\
a_7 & a_{49} & a_{77}\ldots\ldots\ldots\ldots & & & \\
\ldots\ldots\ldots\ldots & & & & &
\end{array}$$

Comparing this arrangement with the standard scheme for letters with double suffixes p, q :—

$$\begin{array}{lll}
b_{11} & b_{12} & b_{13}\ldots\ldots \\
b_{21} & b_{22} & b_{23}\ldots\ldots \\
b_{31} & b_{32} & b_{33}\ldots\ldots \\
\ldots\ldots\ldots\ldots & &
\end{array}$$

we have the numbers 1, 2, 3, ... in a 1, 1 correspondence with the pairs to make room (p, q) in which p, q take, independently of one another, all positive integral values; namely, 1, 2, 3, 4, 5, 6, ... correspond to (1, 1), (1, 2), (2, 1), (1, 3), (3, 1), (1, 4), ...

A simpler arrangement for single suffixes is afforded by

$$\begin{array}{lllll}
a_1 & a_2 & a_4 & a_7 & a_{11}\ldots\ldots \\
a_3 & a_5 & a_8 & a_{12}\ldots\ldots & \\
a_6 & a_9 & a_{13}\ldots\ldots & & \\
a_{10} & a_{14}\ldots\ldots & & & \\
a_{15}\ldots\ldots & & & &
\end{array}$$

In this case $a_1, a_2, a_3, a_4, a_5, a_6, \ldots$ can be written $b_{11}, b_{12}, b_{21}, b_{13}, b_{22}, b_{31}, \ldots$. *It is important to notice that the* 1, 1 *correspondence provides against omissions or repetitions of terms.*

Suppose that a (1, 1) correspondence has been established between the single numbers 1, 2, 3, ... and the pairs (p, q), and that in this way *the terms of an absolutely convergent series* Σa_n have been arranged in rectangular array, the following two questions suggest themselves :—

(1) Is the series formed by each row absolutely convergent?

(2) Granting that such is the case, is the series formed by the sums of the first, second, third, ..., rows absolutely convergent, and is its sum equal to that of the given series?

These questions are answered by the following theorems:—

Theorem I. *When infinitely many terms $b_{m1}, b_{m2}, b_{m3}, \ldots,$ are selected from the absolutely convergent series Σa_n, the series Σb_{mn} is absolutely convergent.*

For the more terms we take the greater is $\Sigma\,|\,b_{mn}\,|$; but it is less than $\Sigma\,|\,a_n\,|$, which by hypothesis is finite. Hence, by the theorem of § 45, $\Sigma\,|\,b_{mn}\,|$ has a limit.

Theorem II. *If by means of a* 1, 1 *correspondence between the numbers n and the pairs (p, q) we arrange the terms of the absolutely convergent series Σa_n in the rows $b_{11}, b_{12}, b_{13}, \ldots$ to infinity, $b_{21}, b_{22}, b_{23}, \ldots$ to infinity, $b_{31}, b_{32}, b_{33}, \ldots$ to infinity, and so on, then the series composed of the sums of these rows is also absolutely convergent and has the same sum as the original series.*

By Theorem I. the series $b_{m1} + b_{m2} + b_{m3} + \ldots$ is convergent; let it have the sum b_m. The series that we have now to consider is $\sum\limits_{m=1}^{\infty} b_m$. Consider the p rows following the mth one.

Let $$r_{m,p} = b_{m+1} + b_{m+2} + b_{m+3} + \ldots + b_{m+p};$$

and let the a with minimum suffix used in the p rows be a_{n+1}. Then $r_{m,p}$ consists of terms selected from r_n, where

$$r_n = a_{n+1} + a_{n+2} + a_{n+3} + \ldots \text{ to infinity},$$

and we infer that

$$|\,r_{m,p}\,| < |\,a_{n+1}\,| + |\,a_{n+2}\,| + |\,a_{n+3}\,| + \ldots.$$

As the expression on the right-hand side tends to zero when n tends to infinity, and as n tends to infinity with m, it follows that $\mathrm{L}r_{m,p} = 0$, independently of the value assigned to p. This establishes the convergence of the series Σb_m.

To complete the proof of the theorem let

$$s_n = a_1 + a_2 + \ldots + a_n, \quad t_m = b_1 + b_2 + \ldots + b_m,$$

and let a_{n+1} be the first of the a's which does not occur in t_m.

The difference $t_m - s_n$ consists of terms a selected from those that occur after a_n, and therefore

$$|t_m - s_n| < |a_{n+1}| + |a_{n+2}| + |a_{n+3}| + \ldots.$$

Hence, when m and n tend to infinity, $|t_m - s_n|$ tends to zero, and we have

$$\mathrm{L}t_m = \mathrm{L}s_n.$$

This proves that the sums of the series Σa_n, Σb_m are equal.

69. Conversion of a Double Series into a Single Series. In the preceding article we have converted an absolutely convergent single series into a double series; let us now examine the reverse process.

Given an array of elements $|b_{pq}|$, *or* B_{pq}, *let the rows yield convergent series with sums* B_1, B_2, B_3, ... *and let the series* ΣB_m *converge also. Then the series formed by adding the sums* b_1, b_2, b_3, ... *of the rows in the array of elements* b_{pq} *will converge absolutely and so also will the single series*

$$b_{11} + b_{12} + b_{21} + b_{13} + b_{22} + b_{31} + \ldots;$$

and these two series will have the same sum.

By the conditions of the theorem each row in the array of b's yields a convergent series; hence the rows have sums b_1, b_2, b_3, ... Since

$$|b_{m1} + b_{m2} + b_{m3} + \ldots| \leqq B_{m1} + B_{m2} + B_{m3} + \ldots,$$

we have $|b_m| \leqq B_m$, and therefore Σb_m is an absolutely convergent series.

Reference to the second array of § 68, shows that

$$b_{11} + b_{12} + b_{21} + b_{13} + b_{22} + b_{31} + \ldots + b_{1h} + b_{2h-1} + \ldots + b_{h1}$$

is the sum of the terms in the first h diagonals. We have to show that the series continued to infinity is absolutely convergent. Consider the rectangle of terms B_{pq} where $p < m + 1$, $q < n + 1$.

Let

$$\begin{aligned}
S &= B_1 + B_2 + \ldots + B_m + R,\\
B_1 &= B_{11} + B_{12} + \ldots + B_{1n} + R_1,\\
B_2 &= B_{21} + B_{22} + \ldots + B_{2n} + R_2,\\
&\vdots\\
B_m &= B_{m1} + B_{m2} + \ldots + B_{mn} + R_m,
\end{aligned}$$

where $R_1, R_2, \ldots, R_m$ can be made less than $\epsilon/2m$ by taking n large enough, and R can be made less than $\epsilon/2$ by taking m large enough. The sum $R+R_1+R_2+\ldots+R_m$ can therefore be made less than ϵ.

$$
\begin{array}{|cccc|c}
\hline
B_{11} & B_{12} & \ldots & B_{1n} & R_1 \\
B_{21} & B_{22} & \ldots & B_{2n} & R_2 \\
\vdots & & & & \vdots \\
B_{m1} & B_{m2} & \ldots & B_{mn} & R_m \\
\hline
\multicolumn{5}{|c}{R}
\end{array}
$$

By taking h large enough we can make the hth diagonal lie wholly without the rectangle, no matter how large m and n may be. This means that the terms, after the kth, of the series

$$B_{11}+B_{12}+B_{21}+\ldots+B_{1h}+B_{2h-1}+\ldots+B_{h1}+\ldots \quad \ldots\ldots(1),$$

can be made to lie wholly without the rectangle by taking k large enough. Hence, if T_k be the sum of k terms of (1),

$$T_{k+l}-T_k<R+R_1+R_2+\ldots+R_m;$$

that is,

$$T_{k+l}-T_k<\epsilon.$$

This proves the convergence of (1) and therefore the absolute convergence of

$$b_{11}+b_{12}+b_{21}+b_{13}+b_{22}+b_{31}+\ldots+b_{1h}+b_{2h-1}+\ldots+b_{h1}+\ldots \quad \ldots\ldots\ldots\ldots\ldots(2).$$

Having proved that the *single* series (2) converges absolutely, Theorem II. of § 68 shows that the array of elements b_{pq} has the same sum whether added by rows or by diagonals.

Corollary. *Under the conditions of the theorem the sum by columns is also equal to the sum by diagonals.*

This also is an immediate consequence of the absolute convergence of the series formed by diagonal summation.

The ordinary modes of summation of the elements b_{pq} of a rectangular array are by rows, columns and diagonals. We

have proved that when the series formed from the elements $|b_{pq}|$ by the first mode is convergent, we can assert that the three modes of summation of the array of elements b_{pq} yield the same sum and that the single series that arises from the third mode is absolutely convergent. We say, in this case, that the double series Σb_{pq} is *absolutely convergent.*

A possible error may be guarded against. With the same notation, we might suppose that we can always add by columns when the rows are absolutely convergent and their sums $b_1, b_2, b_3, \ldots$ form an absolutely convergent series. But consider the double series

$$\begin{aligned} &1 + \xi + \xi^2/2! + \xi^3/3! + \ldots \\ &+ 1 + 2\xi + (2\xi)^2/2! + (2\xi)^3/3! + \ldots \\ &+ 1 + 3\xi + (3\xi)^2/2! + (3\xi)^3/3! + \ldots \\ &+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

Here the rows are convergent for every ξ; their sums are $\exp \xi$, $\exp 2\xi$, ..., and form an absolutely convergent series when $\exp \xi < 1$, that is when $\xi < 0$. But summation by columns is manifestly inadmissible.

It happens frequently that double series proceed both ways; that is, p and q in b_{pq} can take all integer values from $-\infty$ to $+\infty$. We shall see later (§ 128) how such a series is best reduced to a single series.

Ex. Prove that if Σa_n and Σb_n are absolutely convergent then the series

$$a_1b_1 + (a_1b_2 + a_2b_1) + (a_1b_3 + a_2b_2 + a_3b_1) + \ldots$$

is absolutely convergent, and its sum is the product of the sums of the given series.

Deduce the binomial theorem for a negative integral exponent.

CHAPTER IX.

UNIFORM CONVERGENCE OF REAL SERIES.

70. The need of a further classification. Hitherto the general terms of our series have not been considered as dependent on a *variable* x; now we shall make the terms dependent on x. There is clearly a difference between the problem of convergence of say Σx^n for an assigned x and that of the same series when x is treated as a variable; for in the latter case we have to consider, not a single series, but an infinity of series arising from the various values of x in some assigned interval or region.

We consider in this chapter the series

$$f_1\xi + f_2\xi + f_3\xi + \ldots + f_n\xi + \ldots,$$

where the terms are functions of a real variable as defined in ch. VI. § 48; these functions are supposed to be one-valued and continuous for the interval considered. The remainder after n terms will be denoted by $r_n(\xi)$.

The geometric series $\sum\limits_{n=0}^{\infty} \xi^n$ converges only when $|\xi| < 1$; this shows that the convergence of the series is to be considered for intervals,—the open interval $(-1, 1)$ in the present case, and for the general case $\sum\limits_{n=0}^{\infty} f_n\xi$ intervals such as $\alpha \leqq \xi \leqq \beta$, $\alpha < \xi \leqq \beta$, $\alpha \leqq \xi < \beta$, $\alpha < \xi < \beta$, the first and last being closed and open respectively, the others partially open and α, β finite.

One important question is: the terms being one-valued and

continuous functions of ξ in an interval, is their sum a continuous function of ξ in that interval?

Take the simple case of the above geometric series and multiply each term by $1-\xi$. We have then the series

$$(1-\xi)+\xi(1-\xi)+\xi^2(1-\xi)+\ldots,$$

and the sum of n terms is

$$s_n = 1-\xi^n.$$

For an assigned ξ such that

$$0 \leqq \xi \leqq 1,$$

this series is convergent; and it has the sum

$$1 \text{ when } 0 \leqq \xi < 1,$$
$$0 \text{ when } \xi = 1.$$

The sum is therefore not a continuous function of ξ *in the interval of convergence.*

Observe that this peculiarity is due to the double limit;

$$\lim_{n=\infty} \lim_{\xi=1} (1-\xi^n) \text{ is } 0,$$

but

$$\lim_{\xi=1} \lim_{n=\infty} (1-\xi^n) \text{ is } 1.$$

We have then to determine a sufficiently large class of series $\Sigma f_n\xi$ which shall inevitably have continuous functions as their sums. For the extension of the notion of a function of a complex variable must be made by taking some limit; and of the limits open to us that of a series offers the fewest difficulties; and it is highly desirable to know when the function so arrived at has the property of continuity.

A series

$$f_1\xi + f_2\xi + f_3\xi + \ldots + f_n\xi + \ldots,$$

formed of functions which are one-valued and continuous throughout the closed interval (α, β) is, we know, convergent at a point ξ of that interval if $|r_n(\xi)| < \epsilon$ when $n \geqq \mu$.

Let $\xi_1, \xi_2, \ldots, \xi_p$ be p points of the interval, and let $\mu_1, \mu_2, \ldots, \mu_p$ be positive integers,—in the interest of simplicity we shall take them to be the *minimum* positive integers,—for which we can assert that $|r_n(\xi_1)| < \epsilon$ when $n \geqq \mu_1$, $|r_n(\xi_2)| < \epsilon$ when

$n \geqq \mu_2, \ldots, |r_n(\xi_p)| < \epsilon$ when $n \geqq \mu_p$. It is evident that these p inequalities can be condensed into the single inequality

$$|r_n(\xi)| < \epsilon \text{ when } n \geqq \mu,$$

where μ is equal to or greater than the greatest of $\mu_1, \mu_2, \ldots, \mu_p$.

Now when we consider the points of an interval there are infinitely many points and infinitely many μ's to be considered; it is no longer certain that these μ's have a finite upper limit, with the concurrent advantage of a single inequality that will serve for every point ξ of the interval.

These remarks have paved the way for the examination of what is known as uniform convergence.

71. Uniform Convergence. *If for each ϵ we can select a positive integer μ which is independent of ξ and such that*

$$|r_n(\xi)| < \epsilon \textit{ when } n \geqq \mu,$$

whatever be the value of ξ in an interval (α, β), the series is said to converge uniformly in that interval.

In the above case $\sum\limits_{n=0}^{\infty} \xi^{n-1}(1-\xi)$ this can not be done. We have when $-1 < \xi < 1$, $r_n(\xi) = \xi^n$. Now it is true that when ξ is taken arbitrarily and then fixed we can assign a μ such that

$$|\xi^\mu| < \epsilon;$$

for we can always choose μ so that, using Briggs's logarithms,

$$\mu > \frac{\text{Log}_{10} 1/\epsilon}{\text{Log}_{10} 1/|\xi|},$$

when ξ is given. But equally we can always choose ξ so that

$$\mu < \frac{\text{Log}_{10} 1/\epsilon}{\text{Log}_{10} 1/|\xi|},$$

when μ is taken arbitrarily and then fixed.

The point is that the least μ will in general depend on ξ and the values of the μ's for all values of ξ in the interval may or may not have a finite upper limit. Only when this finite upper limit exists is the convergence uniform.

Non-uniform convergence in an interval is due to what is known as *infinitely slow convergence* near certain points of that interval. For an assigned ξ let μ_ξ, or briefly μ, be

the least number of terms of a convergent series which will make $|r_n(\xi)| < \epsilon$ for all values $n \geqq \mu$; then μ is a measure of the rapidity of the convergence. If for example we wish to calculate from a series the value of π with an error $1/10^{6}$ we should say that of two series used for this purpose that one converges the more rapidly which needs the fewest terms to furnish the desired result. Hitherto ξ has been fixed; now let it traverse an interval in which the series converges but not uniformly. As ξ tends to certain values in the interval we have to take more and more terms without limit to secure the desired approximation. The convergence near these points is said to be infinitely slow. Thus in the example $\Sigma\xi^n(1-\xi)$, the convergence is infinitely slow as ξ approaches 1.

It must not be supposed that because μ tends to ∞ as ξ tends to 1, therefore μ is infinite when ξ is 1. In fact when ξ is 1, each term of the series is 0 and therefore

$$|r_n(\xi)| < \epsilon$$

when $n \geqq 1$.

This illustrates the essential distinction between a limit when x tends to a and a value when x is a.

We have defined uniform convergence by means of the remainder $r_n(\xi)$. We could use instead the partial remainder $s_{n+p}(\xi) - s_n(\xi)$ where $s_n(\xi)$ is the sum of the first n terms. In fact if

$$|r_n(\xi)| < \epsilon/2, \ |r_{n+p}(\xi)| < \epsilon/2, \ n \geqq \mu,$$

then
$$|r_{n+p}(\xi) - r_n(\xi)| < \epsilon, \ n \geqq \mu,$$

that is,
$$|f_{n+1}\xi + f_{n+2}\xi + \ldots + f_{n+p}\xi| < \epsilon,$$
$$n \geqq \mu, p = 1, 2, 3, \ldots,$$

μ as before having one and the same value for all points of the interval.

Uniformly convergent series afford a very good example of uniform convergence to a limit (§ 55). Instead of saying that the series converges uniformly we can say equally well that $s_n(\xi)$ converges uniformly to the limit $s(\xi)$, or that $r_n(\xi)$ converges uniformly to the limit 0.

Some writers replace the words uniformly convergent by *equably convergent* or *equiconvergent*; no doubt because the word uniform is used in several senses in pure and applied mathematics.

Ex. 1. Is the series

$$(1-\xi)+\xi(1-\xi)+\xi^2(1-\xi)+\dots$$

uniformly convergent in the interval $-1 \leqq \xi \leqq 0$?

Ex. 2. Prove that the series

$$\frac{\xi}{\xi+1}+\frac{\xi}{(\xi+1)(2\xi+1)}+\frac{\xi}{(2\xi+1)(3\xi+1)}+\dots$$

is not uniformly convergent in an interval which contains 0.

72. Uniform Convergence implies Continuity.

We prove now that *when the series $\Sigma f_n\xi$ is uniformly convergent at all points of the closed interval* (α, β) *the sum of the series is a continuous function of ξ in the interval.*

Because the convergence is uniform there exists a positive integer μ such that

$$|r_n(\xi)| < \epsilon/3 \text{ when } n \geqq \mu,$$

and in particular $\quad |r_\mu(\xi)| < \epsilon/3 \quad\dots\dots(1),$

whatever be the position of ξ in the interval.

Hence if ξ' be another point of the interval

$$|r_\mu(\xi')| < \epsilon/3 \quad\dots\dots(2).$$

Also the sum of the first μ terms, $s_\mu(\xi)$, is the sum of a definite number of continuous functions, and therefore is itself continuous. Therefore it is possible to choose $\xi'-\xi$ so small that

$$|s_\mu(\xi')-s_\mu(\xi)| < \epsilon/3 \quad\dots\dots(3).$$

From (1), (2), and (3),

$$|s_\mu(\xi')-s_\mu(\xi)+r_\mu(\xi')-r_\mu(\xi)| < \epsilon,$$

or if $f\xi = s_\mu(\xi)+r_\mu(\xi)$ = the sum of the series,

$$|f\xi'-f\xi| < \epsilon.$$

And this is the criterion of continuity.

Corollary. *It follows that if the sum of a series of continuous functions is discontinuous, the convergence is not uniform.*

The converse theorem that when the sum of a series is continuous, the convergence is uniform, is not true. For example take the series

$$\left(\frac{\xi}{\xi^2+1}-\frac{2\xi}{2^2\xi^2+1}\right)+\left(\frac{2\xi}{2^2\xi^2+1}-\frac{3\xi}{3^2\xi^2+1}\right)+\dots.$$

Here
$$s_n = \frac{\xi}{\xi^2+1} - \frac{(n+1)\xi}{(n+1)^2\xi^2+1},$$
$$\mathrm{L}\, s_n = \frac{\xi}{\xi^2+1}, \quad \xi \neq 0,$$
$$\mathrm{L}\, s_n = 0, \qquad \xi = 0,$$
and $\frac{\xi}{\xi^2+1}$ is continuous at $\xi = 0$ and takes there the value 0 But at the same time the convergence is not uniform in any closed interval which contains $\xi = 0$.

For
$$r_\mu(\xi) = \frac{(\mu+1)\xi}{(\mu+1)^2\xi^2+1};$$
and in order that $|r_n(\xi)|$ may be $< \epsilon/2$ when $n \geqq \mu$, we must have
$$(\mu+1)|\xi| > 1/\epsilon + (1/\epsilon^2 - 1)^{1/2},$$
where we suppose $\epsilon < 1$. Here as ξ tends to 0, μ tends to ∞.

Ex. Prove that if a series is uniformly convergent in an open interval (a, β) it is convergent at a and β.

73. Uniform and Absolute Convergence.

It must be noticed that uniform convergence does not imply absolute convergence, nor conversely.

Thus the non-uniformly convergent real series
$$(1-\xi) + \xi(1-\xi) + \xi^2(1-\xi) + \dots$$
is absolutely convergent in the interval $0 \leqq \xi \leqq 1$.

On the other hand take any uniformly convergent series. If it is absolutely convergent we can at once turn it into a series conditionally convergent throughout the interval considered, without altering the sum or the uniformity of the convergence, by adding to each term the corresponding term of the series
$$1 - 1 + 1/2 - 1/2 + 1/3 - 1/3 + \dots$$
which converges conditionally to zero*.

A sufficient but not a necessary test for the coexistence of these two kinds of convergence is given by the following theorem :—

* See a paper by Osgood, *Bull. Amer. Math. Soc.* 2nd ser. vol. iii. p. 73. This paper will be found valuable in the way of clearing up many of the difficulties of the subjects of limits and convergence

The series $\Sigma f_n \xi$ is absolutely and uniformly convergent in an interval when the absolute values of its terms for that interval are less than the corresponding terms of a given convergent series $\Sigma \alpha_n$ whose terms are real and positive.

The series converges absolutely by § 64. But further the two inequalities

$$\alpha_{n+1} + \alpha_{n+2} + \ldots < \epsilon \qquad (n \geqq \mu),$$

and

$$|r_n(\xi)| \leqq |f_{n+1}| + |f_{n+2}| + \ldots$$
$$< \alpha_{n+1} + \alpha_{n+2} + \ldots,$$

show that $|r_n(\xi)| < \epsilon$, *whatever be the position of ξ in the interval.* This establishes the fact that the convergence is uniform in the interval in question.

Ex. 1. Prove that every series obtained by multiplying the terms of an absolutely convergent series Σa_n by functions $f_1, f_2, f_3, \ldots$ of the variable ξ which have finite values within an assigned interval, converges absolutely and uniformly throughout that interval.

Prove in particular that

$$\cos \lambda_0 \xi - \tfrac{1}{2} \cos \lambda_1 \xi + \tfrac{1}{4} \cos \lambda_2 \xi - \tfrac{1}{8} \cos \lambda_3 \xi + \ldots$$

is absolutely and uniformly convergent in every closed interval of the axis of real numbers.

Ex. 2. Prove that $\Sigma (-1)^{n-1} \xi^{n-1}/n^2$ is absolutely and uniformly convergent in the closed interval $(-1, 1)$.

Ex. 3. Prove that the series

$$\frac{1+x}{1-x} + \frac{2x}{x^2-1} + \frac{2x^2}{x^4-1} + \ldots + \frac{2x^{2^{n-1}}}{x^{2^n}-1} + \ldots$$

is uniformly convergent along a circle whose centre is 0 and radius < 1.

74. The real Power Series. We shall discuss the case of a power series, that is, a series of ascending positive integral powers,

$$\alpha_0 + \alpha_1 \xi + \alpha_2 \xi^2 + \ldots .$$

Suppose that it converges absolutely for a given value ξ_0. Then by the theorem of § 73 it converges absolutely and uniformly for every ξ such that $|\xi| < |\xi_0|$.

So far as absolute convergence of the above series is concerned, the following theorem gives important additional information.

If for a given value ξ_0 of ξ we have, for every n,

$$|a_n \xi_0^n| \leqq \gamma,$$

where γ is a given positive number, then the series is absolutely convergent for all values of ξ such that $|\xi| < |\xi_0|$.

For
$$|a_n \xi^n| < \gamma\, |\xi/\xi_0|^n,$$
whence the series of absolute values is less than the geometric series
$$\gamma(1 + |\xi/\xi_0| + |\xi/\xi_0|^2 + \ldots),$$
whose sum is
$$\gamma/(1 - |\xi/\xi_0|).$$

The additional information is this: that if the series converges, absolutely or not, for $\xi = \xi_0$, it converges absolutely for $|\xi| < |\xi_0|$. For we can extract from the known fact of convergence at ξ_0 an inequality
$$|a_n \xi_0^n| < \gamma.$$
If this be not obvious, observe that the inequality
$$|a_n \xi_0^n| < \epsilon_0 \qquad n \geqq \mu,$$
where ϵ_0 is any assigned positive number, combined with the existence of a maximum value among the μ terms
$$|a_0|,\ |a_1 \xi_0|,\ \ldots,\ |a_{\mu-1} \xi_0^{\mu-1}|,$$
implies such an inequality provided that $\gamma > \epsilon_0$ and $\gamma >$ the maximum value just mentioned.

It follows that *if the series diverges when $\xi = \xi_1$, it diverges when $|\xi| > |\xi_1|$.* For if it converges for a value of ξ such that $|\xi|$ is greater than $|\xi_1|$ it must converge for ξ_1 itself.

There must then be a positive number ρ such that $\Sigma a_n \xi^n$ converges absolutely when $|\xi| < \rho$ and diverges when $|\xi| > \rho$. We know then the interval of absolute convergence of the series (1), namely it is the interval $(-\rho, \rho)$; but the interval will be closed or open, that is, it will or will not contain the frontier points ρ and $-\rho$, according to circumstances.

We know, so far, that $(-\rho+\beta,\ \rho-\beta)$ will serve as an interval of uniform convergence; here β is any assigned positive number less than ρ. The notion of uniform convergence once firmly grasped it will be evident that it does not matter whether we make $(-\rho+\beta,\ \rho-\beta)$ open or closed or partially open; for

uniform convergence in any one case is accompanied by uniform convergence in the others, since there is convergence at the ends. We have now to examine the more extended interval $(-\rho, \rho)$.

Now if the series is divergent at ρ and at $-\rho$ we have to leave the interval as it stands; it cannot come up to the point ρ, that is, nearer than any assigned distance β, for (§ 72) uniform convergence up to ρ would imply continuity of the sum at ρ. *But if the series be convergent at ρ the interval does come up to ρ.* Take first a special case, the logarithmic series

$$\xi - \xi^2/2 + \xi^3/3 - \dots \qquad \dots\dots\dots\dots(1).$$

That $\rho = 1$ appears from the fact that when $\xi = 1$ we have the conditionally convergent series

$$\operatorname{Log} 2 = 1 - 1/2 + 1/3 - 1/4 + \dots.$$

Let s_n denote the sum of the first n terms of this series. Then the coefficients of (1) are s_1, $s_2 - s_1$, $s_3 - s_2$, ..., and the sum of the terms after the nth is

$$\begin{aligned} r_n(\xi) &= (s_{n+1} - s_n)\,\xi^{n+1} + (s_{n+2} - s_{n+1})\,\xi^{n+2} + \dots \\ &= -s_n\,\xi^{n+1} + \xi^{n+1}(1 - \xi)(s_{n+1} + s_{n+2}\,\xi + s_{n+3}\,\xi^2 + \dots), \end{aligned}$$

or, for every ξ such that $|\xi| < 1$,

$$r_n(\xi) = \xi^{n+1}(1 - \xi)\,[s_{n+1} - s_n + (s_{n+2} - s_n)\,\xi + (s_{n+3} - s_n)\,\xi^2 + \dots] \qquad \dots\dots\dots\dots(2).$$

Now we know that $|s_{n+p} - s_n| < \epsilon$ when $n \geqq \mu$; hence the absolute value of the expression (2), when $0 < \xi < 1$, is less than

$$\xi^{n+1}(1 - \xi)(\epsilon + \epsilon\xi + \epsilon\xi^2 + \dots),$$

that is,

$$< \epsilon\xi^{n+1} < \epsilon.$$

But so soon as ξ disappears we recognize that the convergence is uniform.

We have proved then that (1) is uniformly convergent when $0 < \xi < 1$. Therefore also it is uniformly convergent when $0 \leqq \xi \leqq 1$.

That the interval can be extended to the point 1 is due to the convergence when ξ is 1.

The above argument applies to any series of *ascending* positive integral powers of ξ which is convergent, absolutely or

not, when $\xi=\rho$. For we have used no special characteristic of the logarithmic series, but only the condition of convergence when $\xi=1$. The same argument then applies to the series

$$\alpha_0+\alpha_1\xi+\alpha_2\xi^2+\dots,$$

where $\alpha_0+\alpha_1\rho+\alpha_2\rho^2+\dots$ has a limit s,

provided we now understand by s_n the sum of n terms of this last series.

To sum the matter up, let ρ and $-\rho$ be the frontier points of the series $\Sigma\alpha_n\xi^n$. The interval of convergence is at least $-\rho<\xi<\rho$; the series may or may not converge when $\xi=\pm\rho$. The interval of absolute convergence is at least $-\rho<\xi<\rho$; the series may or may not converge absolutely when $\xi=\pm\rho$. The interval of uniform convergence is at least $-\rho+\beta\leqq\xi\leqq\rho-\beta$, where β is any assigned positive number $<\rho$; but if the series is convergent at ρ (or $-\rho$), the interval extends up to ρ (or $-\rho$). A closed interval of convergence is an interval of uniform convergence.

Why we have to lay stress on the fact that the series of powers is taken in the natural or ascending order will be clear if we consider once more the logarithmic case.

If we rearrange (1) as follows,

$$\xi+\xi^3/3-\xi^2/2+\xi^5/5+\xi^7/7-\xi^4/4+\dots:$$

we make no alteration in the sum so long as $|\xi|<1$, but the sum when $\xi=1$ is now $\frac{3}{2}$ Log 2 (§ 65). Hence the sum of the rearranged series is not continuous, and the rearranged series is not uniformly convergent up to the point 1, but only up to any assigned point $1-\beta$. The limit of the rearranged series when ξ tends, from the left, to 1 is still Log 2; but the value at 1 is $\frac{3}{2}$ Log 2.

Ex. 1. Prove that $\sum\limits_{n=1}^{\infty}\xi^n/n^2$ is absolutely and uniformly convergent when $-1\leqq\xi\leqq1$.

Ex. 2. Prove that $\sum\limits_{n=1}^{\infty}n\xi^{n-1}$ is absolutely convergent when $-1<\xi<1$, and uniformly convergent when $-1+\beta\leqq\xi\leqq1-\beta$, where β is a positive proper fraction.

Ex. 3. The series $\xi^2-\xi^4/2+\xi^6/3-\xi^8/4+\dots$ is uniformly convergent in the interval $-1\leqq\xi\leqq1$.

CHAPTER X.

POWER SERIES.

75. Notation. The series

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots$$

whose coefficients a are independent of x, is called a *power series* or *integral series* in x. Power series play a very important part both in pure and applied mathematics. An elementary illustration of their utility is afforded by the way in which we write numbers. To express an integer we arrange it in powers of 10. To express a proper fraction we often arrange it in powers of 1/10, and in so doing we are led to the consideration of infinite series. To express an improper fraction we combine two power series,—a series in powers of 10 and a series in powers of 1/10. The idea that series of powers are as serviceable for algebra as for arithmetic was first worked out by Newton*, and in the theory of functions of a complex variable, as it now stands, the theory of such series is the solid foundation for the whole structure.

In this chapter we shall explain the principal properties of power series. It will appear that when x is restricted to lie within a certain circle, the infinite series may be handled in precisely the same way as the terminating power series or polynomial.

* Introduction to the *Geometria Analytica*; see Brill and Noether, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. iii. p. 116.

By means of suitable notations the discussion can be shortened considerably. The absolute values of the variable x and of the coefficients a_n will be represented by capital letters X and A_n. The notation (c, R) will denote by itself the circle whose centre is c and radius is R; in the particular case when $c = 0$ the notation $(0, R)$ can be replaced by (R). The system of all the points in the circle (c, R) will be called the *open* region (c, R); the system of points in and on the circle will be called the *closed* region (c, R). When the points of the open region are supplemented by some but not all of the points of (c, R) the system may be called a partially closed region (c, R).

In cases where the investigation does not hinge upon special values of the coefficients, it is extremely convenient to employ the symbol Px, as a short expression for the series

$$a_0 + a_1 x + a_2 x^2 + \dots.$$

Similarly $P(x-c)$ stands for any series of the form

$$a_0 + a_1(x-c) + a_2(x-c)^2 + \dots.$$

We speak of $P(x-c)$ as a series *about* the point c.

When the series is convergent the symbol signifies the sum of the series; also when there is no risk of confusion the symbol can be used for the series itself, whether convergent or not. The context will indicate in every case the sense which is to be attached to the symbol. As the symbol $P(x-c)$ is used generically for a power series, there is often no objection to the use, in one and the same paragraph, of the single symbol for different series. When it is important to distinguish the various series we use new letters such as Q, or we keep P and use suffixes.

The series $P(1/x)$ is of the form

$$a_0 + a_1/x + a_2/x^2 + \dots + a_n/x^n + \dots;$$

this is a series whose variable terms vanish when $x = \infty$ just as the variable terms of $P(x-c)$ vanish when $x = c$. To make statements about $P(x-c)$ hold for all values of x_1 we agree to regard $P(x-\infty)$ as meaning $P(1/x)$. We speak of $P(1/x)$ as the series about ∞.

Later on we shall have to consider series in both ascending and descending powers, such as

$$a_0 + a_1(x-c) + a_2(x-c)^2 + \dots$$
$$+ a_{-1}(x-c)^{-1} + a_{-2}(x-c)^{-2} + \dots.$$

These need no new notation, as they can be expressed by

$$P(x-c) + P(1/(x-c)).$$

The two constitute a *Laurent series* (§ 122).

We shall also have to consider such series as

$$a_0 + a_1\sqrt{x} + a_2(\sqrt{x})^2 + \dots.$$

This is a power series in $\sqrt{x}$, and is denoted by $P\sqrt{x}$.

It must be emphasized that in speaking of a power series we restrict the word power to mean *positive integer power*; and that with the notation now explained there is no inconvenience in so doing. For other series a qualifying adjective can be added; for instance $P(1/x)$ is a *negative power series* in x, $P\sqrt{x}$ a *fractional power series*.

Lastly it is often convenient to designate by a suffix the first power which actually occurs in the series. Thus we often mean by $P_0 x$ a series in which a_0 is not zero, by $P_1 x$ a series in which a_0 is zero but a_1 is not, and so on.

76. The Circle of Convergence. In § 74 we had the theorem: if a real series of powers is convergent when $\xi = \xi_0$ it is absolutely convergent when $|\xi| < |\xi_0|$. The theorem and its demonstration apply equally to the case of complex power series, i.e. when both coefficients and variable are complex. We have then the important theorem:—

If Px is convergent when $x = x_0$ it is absolutely convergent when $|x| < |x_0|$.

So also, as in § 74, if Px is divergent when $x = x_1$ it is divergent when $|x| > |x_1|$.

Hence there is a frontier value R such that when $|x| < R$ there is absolute convergence but when $|x| > R$ there is divergence. That is, within the circle (R) the series is absolutely convergent and without the circle it is divergent.

The circle (R) is called the *circle of convergence*. The open

region (R) is called the *domain* of the series. This domain supplemented by those points of the circle at which the series is convergent gives the region of convergence.

With regard to R we must mention the two extreme cases $R=0$, $R=\infty$. Series for which $R=0$,—i.e. which converge only when $x=0$,—are to be cast aside as useless; series for which $R=\infty$ are of great importance; they will be considered in detail later on in this book.

The following theorem gives useful information relative to the convergence of Px.

If for some positive number g and for a certain value x_0 of x we have $A_n X_0^n < g$ for all values of n, then Px converges absolutely within (X_0).

The proof is exactly along the lines of § 74 and therefore need not be given here.

Ex. Prove by means of this theorem that if Px converge for x_0 it must converge within (X_0).

The value of the radius of convergence must depend on the coefficients. In the following case the dependence is of a very simple character.

When the coefficients of Px are such that $\mathrm{L}\, A_{n+1}/A_n$ exists and is equal to $1/R$, the radius of convergence of the series is R.

Under the conditions of the theorem the ratio $A_{n+1}X^{n+1}/A_nX^n$ tends to a limit X/R. But the series diverges when this limit is greater than 1 and converges when it is less than 1 (§ 64). Hence (R) is the circle of convergence.

It is worthy of remark that though many important *special* series obey the conditions of this theorem, yet it is not permissible to assume in the *general* case that A_{n+1}/A_n tends to a limit; $1+2x+x^2+2x^3+x^4+2x^5+\ldots$ will serve as an example.

We need not discuss the convergence of a power series on the circle of convergence, as it will not affect the future argument. This is why the case $\mathrm{L}\, a_{n+1}/a_n=1$ in § 64 was dismissed without discussion, though, algebraically, series which present this peculiarity are of some importance. However instances of what can occur should be given.

When a series is absolutely convergent at one point of its circle of convergence, it is so at all points of the circle.

For instance $\Sigma x^n/n^2$ has the circle $|x| = 1$ for its circle of convergence; and since $\Sigma 1/n^2$ is convergent the series $\Sigma x^n/n^2$ is absolutely convergent at all points of the circle.

The series Σx^n is divergent when $|x| = 1$; for the condition $L|x^n| = 0$ is not satisfied. Thus this series is divergent at all points of its circle of convergence.

The series $\Sigma x^n/n$ is divergent when $x = 1$ and conditionally convergent when $x = -1$. Let us consider what happens at other points of the circle $|x| = 1$.

We have, if $\quad s_n = x + x^2/2 + \ldots + x^n/n,$

$$(1 - x) s_n = x - (1 - \tfrac{1}{2}) x^2 - (\tfrac{1}{2} - \tfrac{1}{3}) x^3 - \ldots - \left(\frac{1}{n-1} - \frac{1}{n}\right) x^n - \frac{x^{n+1}}{n}.$$

Hence

$$(1 - x)(s_{n+p} - s_n) = -\left(\frac{1}{n} - \frac{1}{n+1}\right) x^{n+1} - \ldots$$
$$- \left(\frac{1}{n+p-1} - \frac{1}{n+p}\right) x^{n+p} - \frac{x^{n+p+1}}{n+p} + \frac{x^{n+1}}{n}:$$

and when $|x| = 1$

$$|1 - x| \, |s_{n+p} - s_n| < \left(\frac{1}{n} - \frac{1}{n+1}\right) + \ldots$$
$$+ \left(\frac{1}{n+p-1} - \frac{1}{n+p}\right) + \frac{1}{n+p} + \frac{1}{n}$$
$$< \epsilon \quad \text{when } n > 2/\epsilon.$$

Hence when $|x| = 1$ and $x \neq 1$, $|s_{n+p} - s_n|$ satisfies the original criterion for convergence.

But since $\Sigma 1/n$ is divergent, the convergence is only conditional.

Hence *the logarithmic series $\Sigma x^n/n$ is conditionally convergent at all points of its circle of convergence except at $x = 1$, where it is divergent.*

Ex. The series

$$x + x^2/2 + x^3/3 + x^4/4 + x^5/5 + x^6/6 + \ldots,$$

and $$x + x^3/3 + x^2/2 + x^5/5 + x^7/7 + x^4/4 + \ldots,$$

have not the same sum when $x = -1$. Prove that, when $|x| = 1$ but x is neither 1 nor -1, the two series have the same sum.

The remarks of this article apply equally to a series $P(x-c)$, which has a circle of convergence (c, R). Also to a series $P(1/x)$, where however the series is convergent outside its circle of convergence. For let $1/x=y$; the series $P(1/x)$ is the same as the series Py and the sums are the same. If Py is convergent when $|y|<R$, then $P(1/x)$ is convergent when $|x|>1/R$.

77. Uniform Convergence of Complex Series. In speaking of a power series Px we understand, unless the contrary is specified, that the x is a point within the circle of convergence. Also Px will be often used to denote the sum of the series, where that number exists; to repeat an earlier remark, the same symbol denotes very conveniently, both a definitive process and the number that results from that process.

Let $f_1x, f_2x, \ldots, f_nx, \ldots$ be functions (defined or to be defined) which are one-valued and continuous in a region Γ of the x-plane; and let the series

$$f_1x+f_2x+\ldots+f_nx+\ldots$$

converge at all points of Γ. Then if $s_n(x)$ be the sum of the first n terms we can, for an assigned x, find a positive integer μ such that

$$|s_{n+p}(x)-s_n(x)|<\epsilon,$$

where $p=1, 2, 3, \ldots$ and $n \geqq \mu$.

It may be that when ϵ is assigned arbitrarily we can determine a μ which is independent of x and which will therefore serve for all points of the region; the convergence is then said to be *uniform* in that region.

The only difference from § 71 is that now we have a region of the x-plane to consider instead of an interval of the ξ-axis. The proof that the sum of the series is continuous in the region applies word for word. And we still have the test of § 73, namely that the sum is both absolutely and uniformly convergent in a region Γ, when for all points of Γ we have

$$|f_nx|<a_n,$$

where Σa_n is a convergent series of positive terms.

In the case of the power series the region Γ is the aggregate of points within the circle (R); the functions are of the form $a_n x^n$, and are continuous within any circle, and in particular in the region (R).

For the series of α's we may take the terms of the series $\Sigma A_n X_0^n$, where $X_0 = |x_0| < R$.

Hence *a power series is uniformly convergent within and on any circle* (X_0), *where* X_0 *is any assigned number* $< R$; *and therefore the sum of a power series is continuous in any assigned circle concentric with but smaller than the circle of convergence.*

This implies that the sum is continuous at every point of the open region (R). For so soon as a point x is selected where $|x| < R$, we can choose the number X_0 so that $|x| < X_0 < R$, and the sum is continuous in the circle (X_0).

Continuity in the whole of a region implies continuity in a part of that region. Thus the sum is continuous for the system of points on any circle (X) where $X < R$.

Naturally the remarks of this article apply equally to the series $P(x-c)$; also to the series $P(1/x)$ outside its circle of convergence.

78. Cauchy's Theorem on the Coefficients of a Power Series. The theorem in question is not intrinsically important; it derives its importance from the fact that it can be used with effect in the discussion of the functions of a complex variable as defined by power series. It can be enunciated as follows:

Given that Px *converges on a circle* (ρ), *where* ρ *is less than the radius of convergence, and that* g *is the maximum value of* $|Px|$ *on this circle, then* $A_n \leqq g/\rho^n$.

The existence of g follows from the theorem of § 51, for $|Px|$ is continuous on (ρ). The following lemma is required in the proof of the general theorem:

Lemma. *There exists a number* θ, *of absolute value* 1, *which satisfies none of the equations* $x^{m_1} = 1$, $x^{m_2} = 1$, ..., $x^{m_n} = 1$, *where* $m_1, m_2, \ldots, m_n$ *are positive or negative integers.*

The proof is very simple. The numbers of roots of these

equations are $|m_1|, |m_2|, \ldots, |m_n|$. Hence from the *infinitely* many values x for which $|x| = 1$ it is possible to select a value $x = \theta$ which shall differ from the roots of these equations; for the number of these roots is not greater than the number $|m_1| + |m_2| + \ldots + |m_n|$.

Let us divide the proof of Cauchy's theorem into two parts.

I. The first part is concerned with the proof of the following auxiliary theorem

Let $fx = b_0 + b_1x^{m_1} + b_2x^{m_2} + \ldots + b_rx^{m_r}$, *where the exponents are positive or negative integers, then* $|b_0| \leqq g$, *where* g *is the maximum value of* $|fx|$ *on the circle* (ρ).

Let θ be chosen as in the lemma; construct the function

$$\phi(x, n) \equiv \frac{f(x) + f(\theta x) + f(\theta^2 x) + \ldots + f(\theta^{n-1}x)}{n},$$

where n is a positive integer. This function is equal to

$$\frac{1}{n}[nb_0 + b_1x^{m_1}(1 + \theta^{m_1} + \theta^{2m_1} + \ldots \text{ to } n \text{ terms}) + \ldots$$
$$+ b_rx^{m_r}(1 + \theta^{m_r} + \theta^{2m_r} + \ldots \text{ to } n \text{ terms})],$$

that is, to

$$b_0 + \frac{b_1x^{m_1}}{n}\frac{\theta^{nm_1} - 1}{\theta^{m_1} - 1} + \frac{b_2x^{m_2}}{n}\frac{\theta^{nm_2} - 1}{\theta^{m_2} - 1} + \ldots + \frac{b_rx^{m_r}}{n}\frac{\theta^{nm_r} - 1}{\theta^{m_r} - 1}.$$

The absolute values of θ^{nm_1}, θ^{nm_2}, etc., are equal to 1; consequently the second factors in all the terms after the first are finite, for their denominators are finite and different from zero and their numerators do not exceed 2 in absolute value. Hence when n tends to infinity, $\phi(x, n)$ tends to the limit b_0.

Since each of the terms $f(x)$, $f(\theta x)$, ... is less than g in absolute value when x lies on (ρ), it follows that $|\phi| \leqq g$; therefore $|b_0| \leqq g$.

II. Let us consider now the power series

$$Px = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots.$$

We find, after division by x^n, that

$$x^{-n}Px = (a_0x^{-n} + a_1x^{-n+1} + \ldots + a_n + a_{n+1}x + \ldots + a_{n+h}x^{n+h})$$
$$+ (a_{n+h+1}x^{n+h+1} + a_{n+h+2}x^{n+h+2} + \ldots).$$

The infinite series in the lower line can be made as small as we please in absolute value by taking h sufficiently great, whatever be the position of x on the circle (ρ). Thus

$$|x^{-n}Px| \geqq |\text{the expression in the upper line}| - \epsilon\,;$$

that is,

$$\begin{aligned}\epsilon + |x^{-n}Px| &\geqq |\text{the expression in the upper line}|\\ &\geqq A_n, \text{ by the first part of the proof.}\end{aligned}$$

But the maximum value of $|x^{-n}Px|$, for the points of (ρ), is $\rho^{-n}g$. Hence

$$A_n \leqq g/\rho^n,$$

which is the theorem that we set out to prove.

Notice that g is the maximum value of $|Px|$ taken over a line (a circle), not over a region. When the meaning of g is changed to 'a number greater than the maximum value of $|Px|$ on (ρ),' we have $A_n < g/\rho^n$.

The above theorem has been stated only for the ordinary series; but it can be extended with ease so as to cover the case of a Laurent series $\sum\limits_{-\infty}^{\infty} a_n x^n$.

The domain of convergence is no longer circular but annular, consisting namely of the points exterior to (R_1) the circle of convergence of $\sum\limits_{n=-1}^{-\infty} a_n x^n$ and interior to (R_2) the circle of convergence of $\sum\limits_{n=0}^{\infty} a_n x^n$, R_1 being supposed less than R_2. The number ρ is chosen so that $R_1 < \rho < R_2$; and then Cauchy's theorem, as generalized, affirms that $A_n \leqq g/\rho^n$, where g is the maximum value of $\sum\limits_{-\infty}^{\infty} a_n x^n$ on (ρ), no matter whether n be zero, positive, or negative.

79. $P_0 x$ does not vanish near $x = 0$. *When $Px \neq 0$ for $x = 0$, there exists an open region X_0 in which Px does not vanish.* For when $x = 0$, the value of Px is a_0 which is not zero; and since Px is a continuous function, there exists a number X_0 such that $|Px - a_0| < A_0$ when $X < X_0$.

But as this gives no information as to X_0, the following

proof, based on Cauchy's theorem, may be added. Take $\rho < R$ and > 0. Then Cauchy's theorem states that $A_n \leqq g/\rho^n$.

Hence for points x in the open region (ρ) we have

$$\begin{aligned} |a_1x + a_2x^2 + \ldots| &\leqq A_1X + A_2X^2 + \ldots \\ &\leqq g(X/\rho + X^2/\rho^2 + \ldots) \\ &\leqq gX/(\rho - X). \end{aligned}$$

Hence if $gX/(\rho - X) < A_0$, that is, if $X < A_0\rho/(g + A_0)$, then $|Px - a_0| < A_0$ and Px cannot vanish.

Thus we have found a value for X_0; it is

$$A_0\rho/(g + A_0),$$

where ρ is any positive number $< R$, and g is the maximum value of Px over the circle (ρ). This is an extension of the first theorem of § 56.

So far we have considered power series which have a constant term; the following theorem is the natural generalization for the case P_nx, where $n > 0$:

When a_0 vanishes and $a_1, a_2, \ldots$ do not all vanish, there is an open region (X_0) within which Px vanishes at no point other than 0.

For let $Px = x^n(a_n + a_{n+1}x + a_{n+2}x^2 + \ldots)$, where $n > 0$. The series in parentheses has the same circle of convergence as Px; furthermore it has a constant term which does not vanish. Hence by the last theorem there is an open region (X_0) within which the series does not vanish. Evidently, then, this region satisfies the requirements of the theorem enunciated above.

80. Criteria of Identity of Power Series. *If Px vanish within every circle (ρ), where ρ is arbitrarily small, at points distinct from* 0, *the coefficients of Px all vanish.*

This is a direct consequence of the theorems of the preceding article; for they tell us that when the coefficients are not all zero there can be found a circle (ρ) at no point of which, excepting perhaps 0, does Px vanish.

We define a *zero* of Px as a point at which Px vanishes. This definition accords with that of § 61.

The following distributions of zeros require that all the coefficients of Px shall vanish:—

(1) They fill a region which contains 0;

(2) They fill a curve which passes through or terminates at 0;

(3) They form a system $x_1, x_2, x_3, \ldots, x_n, \ldots$ to infinity, such that there are numbers of the system $X_1, X_2, X_3, \ldots, X_n, \ldots$ to infinity which are greater than 0 and less than ϵ. An example is afforded by the system $x_n = (-1)^{n-1}/n$.

In (1) and (2) $a_0 = 0$ because $Px = 0$ when $x = 0$. Let $Px = xQx$; we must not assume that $x = 0$ satisfies $Qx = 0$ and prove that $a_1 = 0$ by putting $x = 0$. The proper proof is the same as that by which we show in (3) that $a_0 = 0$; namely Px cannot be continuous at $x = 0$, if $P0 \neq 0$, while $Px = 0$ at points arbitrarily close to 0.

To meet all cases, we need a new word. When the points of an infinite system are distributed, as in (1), (2), (3), so that one or more points, distinct from x_0, lie within (x_0, ϵ) however small we choose to make ϵ, then the point x_0 is said to be a *limit-point* of the system.

The above theorem can now be stated thus:—

If the zeros of Px have 0 as a limit-point, the coefficients of Px all vanish.

The case where the zeros are infinite in number, but do not have 0 as a limit-point, will be considered in § 88.

The following theorem is an immediate consequence:—

When the equality $\sum\limits_{n=0}^{\infty} a_n x^n = \sum\limits_{n=0}^{\infty} b_n x^n$ is satisfied by values of x which have 0 for a limit-point, we have $a_n = b_n$ for all values of n.

For the zeros of $\Sigma (a_n - b_n) x^n$ have 0 for a limit-point, and therefore $a_n = b_n$ for all values of n.

CHAPTER XI.

OPERATIONS WITH POWER SERIES.

81. Weierstrass's Theorem on Series of Power Series.

Let
$$u_0 + u_1 + u_2 + \ldots + u_q + \ldots$$
be a series of power series with the general term
$$u_q = a_{q0} + a_{q1}x + a_{q2}x^2 + \ldots + a_{qn}x^n + \ldots,$$
and let (R) be a circle within which the separate terms u_q and the collective series Σu_q converge. Arranging the series Σu_q in the form of a rectangular array in which u_q occupies the $(q+1)$th row, it is desirable to know whether the sum by columns is equal to the sum by rows; for the sum by columns is formally a power series in x, and should it be equal to the sum by rows the series of power series can be condensed into a single power series. The following theorem, due to Weierstrass, supplies a *sufficient* condition, different from that in § 69.

Given that the separate series u_q and the collective series Σu_q converge within the circle (R) and that the latter series Σu_q converges uniformly along every circle (R_1), where $R_1 < R$; then within the circle (R) we have
$$\sum_{q=0}^{\infty} u_q = \sum_{n=0}^{\infty} a_n x^n,$$
where a_n is the sum of the coefficients of x^n in the series of u's.

We shall prove successively that the series $\sum_{m=0}^{\infty} a_{mn}$, or the series of coefficients in the $(n+1)$th column, converges: that if the sum of this series be a_n, then $\Sigma a_n x^n$ converges in the circle

(R); and lastly that the series $\Sigma a_n x^n$, Σu_q are equal in this circle.

I. By reason of the uniform convergence of Σu_q along the circle (R_1) we can find a number μ such that for all points x on this circle

$$\left|\sum_{m}^{m+r} u_q\right| < \epsilon, \qquad r = 1, 2, 3, \ldots,$$

as soon as m is chosen equal to or greater than μ. But $\sum_{m}^{m+r} u_q$ is the sum of a *finite* number of convergent series; and therefore, adding by columns, the inequality can be written

$$\left|\sum_{n=0}^{\infty}(a_{mn} + a_{m+1,\,n} + \ldots + a_{m+r,\,n})\, x^n\right| < \epsilon,$$

where $|x| = R_1$. Cauchy's theorem (§ 78) enables us to separate out the coefficient of x^n from the remaining coefficients and gives

$$|a_{mn} + a_{m+1,\,n} + \ldots + a_{m+r,\,n}| < \epsilon/R_1^{\,n},$$

or $$|a_{mn} + a_{m+1,\,n} + \ldots \text{ to infinity}| \leqq \epsilon/R_1^{\,n}.$$

This inequality establishes the convergence of the series $\sum_{m=0}^{\infty} a_{mn}$; it is permissible therefore to assign to this series a sum a_n.

II. Let $a_n = s_{mn} + t_{mn}$, where $m \geqq \mu$ and s_{mn}, t_{mn} are the sum to m terms and the remainder after m terms of the series $\sum_{m=0}^{\infty} a_{mn}$ that arises from the $(n+1)$th column

The sum by columns of the double series can be resolved into two parts, which arise from the columns continued down to the mth row inclusive and from the columns continued from the $(m+1)$th row on to infinity. The sums of these two parts are

$$s_{m0} + s_{m1}x + s_{m2}x^2 + \ldots + s_{mn}x^n + \ldots \ldots\ldots\ldots (1),$$

$$t_{m0} + t_{m1}x + t_{m2}x^2 + \ldots + t_{mn}x^n + \ldots \qquad \ldots\ldots (2).$$

The former of these two series converges in the open region (R), for each of its m component series $u_0, u_1, \ldots, u_{m-1}$ converges in that region. Let us see whether the same is the case for the latter of the two series.

By Cauchy's inequality, we have

$$|t_{mn}| \leqq \epsilon/R_1^n.$$

Hence, using capital letters to denote absolute values,

$$T_{m0} + T_{m1}X + T_{m2}X^2 + \ldots + T_{mn}X^n + \ldots \leqq \epsilon \sum_{n=0}^{\infty} X^n/R_1^n.$$

But when $X < R_1$ the series $\Sigma X^n/R_1^n$ is convergent and has the sum $R_1/(R_1 - X)$; therefore $\sum\limits_{n=0}^{\infty} t_{mn}x^n$ converges absolutely for all values $X < R_1$. This means that it converges absolutely in the open region (R); for, if x be *any* selected point of this region, R_1 can be chosen between X and R and the absolute convergence at x follows at once.

It results from the addition of (1) and (2) that the series $\sum\limits_{n=0}^{\infty} a_n x^n$ converges within the circle (R).

III. Finally we have to prove that

$$\Sigma u_q = \Sigma a_n x^n$$

within the circle (R).

Select any value x within (R) and then take R_1 intermediate between X and R. For this value of x and for the same value of m as before, we have to prove that

$$\sum_{q=0}^{m-1} u_q + \sum_{q=m}^{\infty} u_q = \sum_{n=0}^{\infty} s_{mn}x^n + \sum_{n=0}^{\infty} t_{mn}x^n;$$

or removing the first terms on the two sides of the equation, since these are known to be equal, we have to prove that

$$\sum_{q=m}^{\infty} u_q = \sum_{n=0}^{\infty} t_{mn}x^n \qquad\qquad (3).$$

This can be done very readily by the use of the inequalities that have been found above. For

$$\left|\sum_{q=m}^{\infty} u_q - \sum_{n=0}^{\infty} t_{mn}x^n\right| \leqq \epsilon + \epsilon R_1/(R_1 - X);$$

and whenever it can be asserted of any *fixed* quantity that its absolute value is less than an arbitrarily small positive number, that quantity is necessarily zero; hence (3) is proved.

We proceed to some important applications of the theorem.

82. Remarks on Weierstrass's Theorem. This theorem of Weierstrass's on series of power series in x can be extended at once to series of power series in $1/x$. It must be remembered that now the domain of convergence of a power series is the part of the plane exterior to some circle $(0, R)$. The theorem as extended runs:—*If the terms of the series* Σu_q, *where*

$$u_q = a_{q0} + a_{q1}/x + \ldots + a_{qn}/x^n + \ldots,$$

converge outside the circle (R') *and the series itself converges uniformly on the perimeter of every circle* (R_1') *where* $R_1' > R'$, *then for every point* x *exterior to the circle* (R') *the series* Σu_q *is expressible as a single power series in* $1/x$.

Combining this result with that given earlier we obtain at once a sufficient condition for the composition of a series of series $\sum\limits_{n=-\infty}^{+\infty} a_n x^n$ into a single series of that form. Instead of a region interior to (R) or a region exterior to (R'), we have the annular region (R', R) where R' is supposed less than R.

A convenient modification of Weierstrass's theorem replaces the uniform convergence of the series $\sum\limits_{q=0}^{\infty} u_q$ over every circle (R_1), where $R_1 < R$, by the uniform convergence of the series for the closed regions (R_1). Evidently the former condition is contained in the latter.

We shall, in general, use Weierstrass's criterion for the conversion of the double series Σu_q into the single series Px, in preference to that of Cauchy (§ 69, corollary). There is nothing in either method to indicate whether the region for which the equation is proved to hold good includes all points that satisfy the equation $\Sigma u_q = Px$; later on we shall see that Taylor's theorem furnishes an example where the information furnished by these criteria,—exact of course as far as it goes,—proves to be incomplete.

At first sight Cauchy's criterion seems both simpler and more effective than that of Weierstrass; that this is not always the case appears from the series

$$1 + \frac{x-1}{3-x} + \left(\frac{x-1}{3-x}\right)^2 + \ldots + \left(\frac{x-1}{3-x}\right)^n + \ldots \qquad \ldots\ldots(1),$$

or
$$\sum_{n=0}^{\infty}\left(-\frac{1}{3}+\frac{2}{3^2}x+\frac{2}{3^3}x^2+\dots\right)^n \quad \dots\dots\dots\dots\dots(2).$$

The series (1) converges for all values of x such that
$$\left|\frac{x-1}{3-x}\right|<1,$$
i.e. for all values of x for which the real parts are less than 2; and it converges uniformly for all values of x such that $\left|\frac{x-1}{3-x}\right| \leq \alpha$, where α is a proper fraction. The component series have 3 for a common radius of convergence; hence all the conditions of Weierstrass's theorem are satisfied by taking R to be 2, and we see without further discussion that (2) can be expressed as a power series whose radius of convergence is at least as great as 2.

Cauchy's method uses the *new* series
$$\frac{1}{3}+\frac{2}{3^2}X+\frac{2}{3^3}X^2+\dots=\frac{2}{3}+\frac{X-1}{3-X},$$
replacing $\frac{x-1}{3-x}$ in (1) by $\frac{2}{3}+\frac{X-1}{3-X}$; and requires for the convergence of this new series the inequality
$$\frac{2}{3}+\frac{X-1}{3-X}<1;$$
that is, $X<3/2$. It has given us therefore less information as to the domain of convergence of the final series and has required the summation of a new series.

The two series
$$\frac{x}{1-x}+\frac{x^3}{1-x^3}+\frac{x^5}{1-x^5}+\dots \quad \dots\dots\dots\dots\dots(3),$$
$$\frac{x}{1-x^2}+\frac{x^2}{1-x^4}+\frac{x^3}{1-x^6}+\dots \quad \dots\dots\dots\dots\dots(4),$$
in which $X<1$, can be represented as double series by expanding the individual terms as power series; furthermore they arise from one and the same array
$$\begin{array}{ccccc} x & x^2 & x^3 & x^4 & \dots \\ x^3 & x^6 & x^9 & x^{12} & \dots \\ x^5 & x^{10} & x^{15} & x^{20} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{array}$$

by adding by rows and columns respectively. The double series is absolutely convergent, for the sum by rows of the absolute values of the separate terms is the convergent series

$$\frac{X}{1-X}+\frac{X^3}{1-X^3}+\frac{X^5}{1-X^5}+\dots.$$

Hence Cauchy's criterion applies and the sum by columns must be equal to the sum by rows, and therefore the series (3), (4) have the same sum.

To see that Weierstrass's criterion applies to this case, observe that for any assigned value of x such that $X<1$, we can take R between X and 1 and get

$$\left|\frac{1}{1-x^{2n+1}}\right|<\frac{1}{1-R^{2n+1}}<\frac{1}{1-R};$$

and therefore

$$\left|\sum_{n=0}^{\infty}\frac{x^{2n+1}}{1-x^{2n+1}}\right|<\frac{1}{1-R}\sum_{n=0}^{\infty}\left|x^{2n+1}\right|$$

$\leqq$ a series which is uniformly convergent within (R).

It follows that the series (1) is uniformly convergent within (R) and therefore comes under the operation of Weierstrass's theorem.

Ex. Prove that

$$\frac{x}{1-x}+\frac{2x^2}{1-x^2}+\frac{3x^3}{1-x^3}+\dots=\frac{x}{(1-x)^2}+\frac{x^2}{(1-x^2)^2}+\frac{x^3}{(1-x^3)^2}+\dots$$

when $|x|<1$.

83. Applications of Weierstrass's Theorem. (i) The theorem that *the product of two power series.*

$$Px=\sum_{n=0}^{\infty}a_nx^n,\quad Qx=\sum_{n=0}^{\infty}b_nx^n,$$

is itself a power series affords a simple example of Weierstrass's theorem on double series.

Suppose that Px, Qx both converge in the open region (R), and let us consider the series $\sum\limits_{n=0}^{\infty}f_nx$, where $f_nx=b_nx^nPx$.

This series Σf_nx converges uniformly in the closed region

(R_1) where $0 < R_1 < R$; for it is possible to find a μ such that at every point x of this region (R_1) we have, for values $n \geqq \mu$,

$$|r_n(x)| < \epsilon_1, \quad \epsilon_1 |Px| < \epsilon,$$

where $r_n(x)$ is the remainder of Qx after n terms. We may add therefore by columns and thus get

$$Px \times Qx = \sum_{n=0}^{\infty} (a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \dots + a_n b_0) x^n$$

for all points of the open region (R).

In particular $(Px)^2$ can be arranged as a power series in x which is convergent when Px itself is; and the same can be said of any positive integral power of Px.

Ex. Show that if all three series

$$\Sigma a_n, \quad \Sigma b_n, \quad \Sigma (a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \dots + a_n b_0)$$

converge, then the third is the product of the first two.

(ii) *A sufficient condition that $P(Px)$ shall be expressible as Px.* Suppose that Px is a power series with a radius of convergence R_1 and that $\sum_{n=0}^{\infty} a_n y^n$ is a power series with a radius of convergence R_2; then $\sum_{n=0}^{\infty} a_n (Px)^n$ is not to be treated as a power series in x without a preliminary examination of the circumstances of the case. It is desirable to have a sufficient condition that the transformation $y = Px$ shall convert the series $\sum_{n=0}^{\infty} a_n y^n$ into Px.

First let us simplify matters by making the series in y a terminating power series. We have then a finite number of power series in x associated with $a_1 y$, $a_2 y^2$, $a_3 y^3$ etc. and the sum of a finite number of power series is itself a power series. Thus a polynomial in y will in all cases transform into a power series in x.

But when we pass to the general case of an infinite series in y, we have the sum of an infinite number of power series in x and such a sum is not necessarily expressible as a power series. This case we shall now consider.

It is evident that if $\sum^{\infty} a_n (Px)^n$ is to be represented as a power

series in x we must have the former series convergent for sufficiently small values of x. Thus Px, or y, for values of x very near to zero must be less than R_2; this requires that $|Po|$ must be less than R_2.

Now let this condition, $|Po| < R_2$, be satisfied. Then because Px is continuous there is a circle (R), where $R < R_1$, within which $|Px| < R_2 - \delta$, where δ is a sufficiently small assigned positive number.

Within this circle (R) the series $\Sigma a_n (Px)^n$ is uniformly convergent, since its terms are less than those of the convergent positive series $\Sigma A_n (R_2 - \delta)^n$.

Thus when $|Po| < R_2$ *the conditions of Weierstrass's theorem hold, and therefore it is possible to express the series* $\sum\limits_{n=0}^{\infty} a_n (Px)^n$ *as a power series in* x; *or, symbolically, to express* $P(Px)$ *as* Px.

A case of frequent occurrence is that in which Px in $\Sigma a_n (Px)^n$ begins with a term in x; when this happens the condition of the theorem is evidently satisfied, since $Po = 0$.

Some special cases of this are useful. First, the quotient Px/Qx of two power series, where $Qo \neq 0$, is expressible as a power series. For if $Qx = b_0 + b_1 x + b_2 x^2 + \ldots = b_0 + y$, we have, since $|y| < |b_0|$ for small values of x,

$$\frac{1}{Qx} = \frac{1}{b_0 + y} = \frac{1}{b_0}\left(1 + \frac{y}{b_0}\right)^{-1} = \frac{1}{b_0}\left(1 - \frac{y}{b_0} + \frac{y^2}{b_0^2} - \ldots\right);$$

hence $\dfrac{1}{Qx} =$ a power series in x which converges near $x = 0$

Multiplying each side of the equation by Px and using the theorem of the last article, we see that Px/Qx is expressible as a power series.

When $b_0 = b_1 = \ldots = 0$, and b_m is the first b which does not vanish, the result has to be modified slightly. Now

$$Qx = x^m Q_0 x,$$

and $Px/Qx = Px/x^m Q_0 x$; and since $Px/Q_0 x$ is expressible as a power series, the original series Px/Qx is a Laurent series in which the negative powers stop at x^{-m}.

Again if Px has an infinite radius of convergence,—for

example if $Px = \sum_{n=0}^{\infty} \frac{x^n}{n\,!}$, or $\sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{2n+1\,!}$, or $\sum_{n=0}^{\infty}(-1)^n \frac{x^{2n}}{2n\,!}$,—then $\Sigma a_n (Px)^n$ when converted into a power series in x converges in the domain of $\Sigma a_n y^n$

(iii) If $f(y_1, y_2, \ldots, y_n)$ be a rational integral function in the y's or a power series in the y's which converges for all finite values of those variables, and if the y's be expressible as power series Px converging within a common circle (R), then $f(y_1, y_2, \ldots, y_n)$ can be expressed as a power series which converges within (R).

84 Reversion of a Power Series. Let $y = P_1 x$, and let Qy be expressed as a power series in x. Since $P_1 0$ is 0, the condition in (ii) is merely that Qy shall have a radius of convergence greater than 0. Let us examine whether Qy can be identified with x itself. We have

$$y = a_1 x + a_2 x^2 + \ldots$$

$$x = b_0 + b_1 y + b_2 y^2 + \ldots$$

$$= b_0 + b_1 a_1 x + (b_1 a_2 + b_2 a_1^2)x^2 + (b_1 a_3 + 2b_2 a_1 a_2 + b_3 a_1^3)x^3 + \ldots,$$

so that the conditions to be satisfied are

$$b_0 = 0,$$
$$b_1 a_1 = 1,$$
$$b_1 a_2 + b_2 a_1^2 = 0,$$
$$b_1 a_3 + 2b_2 a_1 a_2 + b_3 a_1^3 = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

and these equations determine $b_0, b_1, b_2, b_3, \ldots$ uniquely.

Hence *if the series $x = P_1 y$ whose coefficients are obtained by the above equations is shown to be convergent for values of y other than* 0, *we have one solution of the equation for x,*

$$y = P_1 x.$$

This very important process is called the *reversion of a power series*; and $P_1 x$ is said to have been *reverted*.

Of course under the same conditions the equation

$$y = a_0 + P_1 x$$

can be reverted in the form

$$x = P_1 (y - a_0).$$

The same process is not applicable to $y = P_n x$ when $n > 1$. For

then $Q(P_n x)$ contains no power of x less than the nth and cannot be identified with x. To this case we shall return later and we shall then be able to establish the convergence of the reverse series. Meanwhile we wish to point out an important peculiarity which arises when we map the domain of $P_n x$ on the y-plane. Taking more generally

$$y - y_0 = P_n(x - x_0),$$

we have first when $n = 1$,

$$y - y_0 = a_1(x - x_0) + a_2(x - x_0)^2 + \dots,$$

and $\lim\limits_{x = x_0}(y - y_0)/(x - x_0) = a_1$, showing that there is isogonality at x_0 (§ 26). But if $a_1 = a_2 = \dots = a_{n-1} = 0$, $a_n \neq 0$, the isogonality breaks down; we have then $\lim\limits_{x=x_0}(y - y_0)/(x - x_0)^n = a_n$, whence an angle at y_0 is n times the corresponding angle at x_0.

Let $\bar{y} - y_0 = a_n(x - x_0)^n$; then $y - \bar{y} = P_{n+p}(x - x_0)$, where p is some positive integer. When x describes a small circle (x_0, ρ), $\bar{y}$ describes a small circle, n times as fast, about y_0. Also

$$(y - \bar{y})/(\bar{y} - y_0) = P_p(x - x_0);$$

hence $|x - x_0|$, or ρ, can be so chosen that $|y - \bar{y}|/|\bar{y} - y_0|$ is as small as we please, for all points on the circle (x_0, ρ). It is usual to say that y describes a small circle about y_0 n times as fast as x describes its circle about x_0; it will be understood that this is strictly true only for $\bar{y}$. The path of y is, for small enough values of ρ, as nearly circular as we please. The case $n = 1$ is the ordinary case; the case $n > 1$ means that when $y = y_0$ there are n values of x equal to x_0. This peculiarity will be discussed in chapters XX. and XXI.

Given that $\qquad y = a_1 x + a_2 x^2 + a_3 x^3 + \dots$

arises from $\qquad x = b_1 y + b_2 y^2 + b_3 y^3 + \dots,$

it is desirable sometimes to have a formula which connects b_n with y. Differentiate (see § 87)

$$x = b_1 y + b_2 y^2 + \dots + b_n y^n + \dots$$

with respect to x and divide by y^n. Then

$$\frac{1}{y^n} = y'\left[\frac{b_1}{y^n} + \frac{2b_2}{y^{n-1}} + \frac{3b_3}{y^{n-2}} + \dots + \frac{nb_n}{y} + (n+1)\,b_{n+1} + \dots\right].$$

The only term on the right-hand side which contains $1/x$ to the

first power is the term $nb_n y'/y$. For example y'/y^n arises, save as to a constant, from $1/y^{n-1}, = \sum_{-(n-1)}^{\infty} c_m x^m$, by differentiation; and after such a series has been differentiated term by term there can be no term in $1/x$ to the first power. By equating the term in $1/x$ in $nb_n y'/y$ to the corresponding term in $1/y^n$, we get

$$b_n = \frac{1}{n}\left[\frac{1}{y^n}\right]_{1/x},$$

where $[\quad]_{1/x}$ means the coefficient of $1/x$ in the expression contained within the square brackets, when that expression is given as a series in x.

Thus the reverse series of

$$y = a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

is

$$x = \sum_{n=1}^{\infty} \frac{y^n}{n}\left[\frac{1}{y^n}\right]_{1/x}$$

85. Taylor's Theorem for Power Series. Let x and $x+h$ lie in the domain (R) of Px. The series for $P(x+h)$ is

$$a_0 + a_1(x+h) + a_2(x+h)^2 + a_3(x+h)^3 + \dots,$$

and may be written

$$a_0 + a_1(x+h) + a_2(x^2 + 2xh + h^2) + a_3(x^3 + 3x^2h + 3xh^2 + h^3) + \dots .$$

Can we arrange this series in powers of h? Treat x as a fixed point within (R), refer h to x as origin, and let u_q (§ 81) be $a_q(x+h)^n$. The series Σu_q itself is $P(x+h)$, and is uniformly convergent along any circle concentric with and smaller than the circle $|x+h| = R$, i.e. for values h within $(x, R-X)$. Take then for the

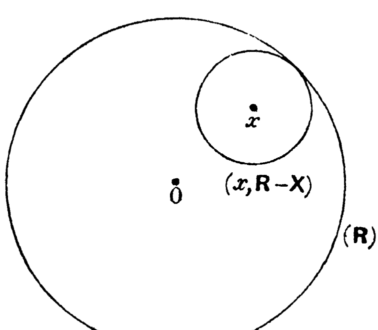


Fig. 39.

circle of Weierstrass's theorem the circle $(x, R - X)$. And suppose, as above, that x lies within (R) and that h is a point interior to the circle $(x, R - X)$ which touches internally the circle of convergence (R). With these limitations on x and h we have

$$P(x+h) = Px + hP'x + \frac{h^2}{2!}P''x + \ldots + \frac{h^n}{n!}P^n x + \ldots,$$

where the series $P'x$, $P''x$, ..., $P^n x$, ..., known as the 1st, 2nd, ..., nth derived series of Px, are

$$P'x = a_1 + 2a_2 x + 3a_3 x^2 + \ldots + (n+1)\,a_{n+1}x^n + \ldots,$$
$$P''x = 1\,.\,2a_2 + 2\,.\,3a_3 x + 3\,.\,4a_4 x^2 + \ldots + (n+2)(n+1)\,a_{n+2}x^n + \ldots,$$

and so on. The series $P'x$ is derived from Px by a differentiation term by term, and a repetition of this process gives successively $P''x$, $P'''x$, etc.

By this use of Weierstrass's theorem we have proved the following theorem which is the exact analogue of Taylor's theorem for real variables:

Let x be any point in the domain (R) of the series Px and let a circle be described with centre x to touch (R) internally; then for all points $x + h$ within this latter circle the series $P(x+h)$ can be represented by the series

$$Px + hP'x + \frac{h^2}{2!}P''x + \ldots + \frac{h^n}{n!}P^n x + \ldots,$$

in which the coefficient of $\frac{h^n}{n!}$ is obtained by differentiating each term of Px n times and adding the results.

Ex. Prove the same by means of § 68.

Replacing x by x_0 and $x + h$ by x, the equation runs

$$Px = Px_0 + P'x_0(x - x_0) + \frac{P''x_0}{2!}(x-x_0)^2 + \ldots + \frac{P^n x_0}{n!}(x-x_0)^n + \ldots.$$

This form of Taylor's series is especially useful in the discussion of power series, for it establishes a connexion between the original series in x and a new series in $x - x_0$, where x_0 lies in the domain of Px.

We now know that the radius of convergence of the series in $x - x_0$ is *at least* as great as $R - X_0$; we shall see presently that

there are many cases in which it can be greater than this quantity.

Maclaurin's theorem

$$Px = P0 + P'0 . x + P''0 . x^2/2! + \ldots + P^n 0 . x^n/n! + \ldots$$

is of course merely a special case of Taylor's theorem.

If we start with a series $P(x - x_0)$ in place of Px, and select a point x_1 within the domain of $P(x - x_0)$ instead of the point x_0 in the domain of Px, Taylor's theorem will run

$$P(x - x_0) = P(x_1 - x_0) + P'(x_1 - x_0) . (x - x_1)$$
$$+ P''(x_1 - x_0) . (x - x_1)^2/2! + \ldots + P^n(x_1 - x_0) . (x - x_1)^n/n! + \ldots$$

The domain of this series in $x - x_1$ is at least as extensive as the open region bounded by that circle of centre x_1 which touches internally the circle of convergence of $P(x - x_0)$.

86. The Derivates of a Power Series. The radius of convergence, R, of $P'x$ will be proved equal to R by establishing that (1) $R' \geqq R$, and (2) $R \geqq R'$.

Since each column in Weierstrass's method for combining a series of power series into a single power series gives a convergent series whatever be the position of x within (R), the series $P'x$ that arises from the first column must give $R' \geqq R$.

To see that $R \geqq R'$ notice that the domain of convergence of Px is the same as that of $\sum_{n=1}^{\infty} a_n x^{n-1}$, and compare this latter series with $P'x$, or $\sum_{n=1}^{\infty} na_n x^{n-1}$. As the general term of the former series is evidently less than that of the latter series in absolute value, the domain of Px is at least as extensive as that of $P'x$. This completes the proof.

It is an immediate deduction that $P''x$, $P'''x, \ldots,$ $P^n x, \ldots$ have a common radius of convergence, namely R. Hence *all the derivates of Px have the same domain as Px.*

Ex. Prove that $\sum_{n=0}^{\infty} a_n x^{n+1}/(n+1)$, $\sum_{n=0}^{\infty} a_n x^{n+2}/(n+1)(n+2), \ldots,$—series obtained by successive uses of the mechanical process of integration term by term,—have a common domain of convergence, namely (R).

Let us now find the value of DPx, the derivate of the sum of the series. At every point of the domain (R) the value of DPx is the limit of $\dfrac{P(x+h)-Px}{h}$ when $x+h$ tends within the domain (R) to x. There is no difficulty in finding this value; for the fraction is equal to a power series in h whose first term is $P'x$, and therefore tends to $P'x$ as a limit. Hence

$$DPx = P'x.$$

This theorem required proof. The existence of $D_\xi f\xi$ is not a necessary consequence of the continuity of $f\xi$ when ξ is real; hence the statement that DPx exists cannot be regarded as a truism. The theorem shows that we get the same result whether we first sum the power series and then differentiate, or first differentiate the separate terms and then sum the resulting series. In the present case then we have a double limit in which the order of operations does not matter.

87. Differentiation of a Series of Power Series term by term. Let $u_0, u_1, u_2, \ldots, u_q, \ldots, R$ have the same meanings as in § 81 and let Σu_q be uniformly convergent within every circle (R_1) where R_1 is less than R by an arbitrarily small amount.

Take any point x_0 within the circle (R) and expand the series $u_0, u_1, u_2, \ldots, u_q, \ldots$ and $u, = \sum\limits_{q=0}^{\infty} u_q$, in ascending powers of $x-x_0$. Then

$$\sum_{m=0}^{\infty} u^{(m)}(x_0)\,.\,(x-x_0)^m/m! = \sum_{q=0}^{\infty}\sum_{m=0}^{\infty} u_q{}^{(m)}(x_0)\,.\,(x-x_0)^m/m! \ldots(1).$$

Since the individual series $u_0, u_1, u_2, \ldots, u_q, \ldots$, and the collective series

$$u = u_0 + u_1 + u_2 + \ldots + u_q + \ldots \ldots\ldots\ldots\ldots\ldots(2),$$

converge uniformly in the closed region $(x_0, R' - X_0)$, we can apply Weierstrass's theorem to the double series in (1) and reduce it to the single series

$$\sum_{m=0}^{\infty} \{u_0{}^{(m)}(x_0) + u_1{}^{(m)}(x_0) + u_2{}^{(m)}(x_0) + \ldots + u_q{}^{(m)}(x_0) + \ldots\}\,(x-x_0)^m/m!\,.$$

Comparing the terms in $x^m/m!$ in this series and in the series on the left-hand side of (1), we deduce that

$$u^{(m)} = u_0{}^{(m)} + u_1{}^{(m)} + u_2{}^{(m)} + \dots + u_q{}^{(m)} + \dots,$$

at every point of the open region (R). In words this theorem runs as follows:—*Under the specified conditions the m^{th} derivate of the sum of a series of power series is equal to the sum of the m^{th} derivates of the individual terms.*

CHAPTER XII.

CONTINUATION OF POWER SERIES.

88. The Zeros of Px are Isolated Points. With the help of Taylor's theorem it is easy to generalize the theorem of § 80. In its generalized form the theorem runs:—

When Px has infinitely many zeros within a circle (R') concentric with and interior to the circle of convergence (R), the coefficients of Px are all zero.

Let the circle be enclosed within the square whose sides

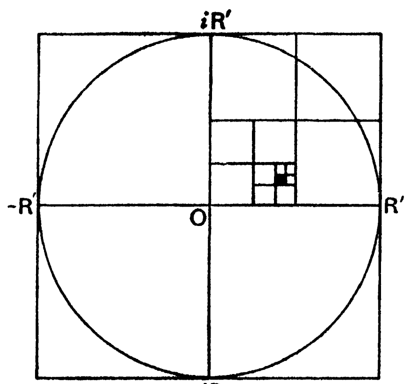


Fig. 40.

are the tangents at $\pm R'$, $\pm iR'$. Divide this square into four smaller squares as in the figure; in at least one of these new squares there must lie infinitely many zeros of Px. Select a square of this kind and subdivide it into four squares as in the figure. Select from these new squares one in which the number

of zeros of Px is infinite and subdivide this square in turn into four new squares. Proceeding in this way we get a sequence of squares, each one of which contains infinitely many zeros of Px. The reasoning of § 51 shows that these squares determine a single point a which is interior to them all. This point is a *limit-point* of the zeros (§ 80), for within *every* circle of centre a, no matter how small may be the radius, there are zeros which are distinct from a. Hence the series $\Sigma P^n a . (x-a)^n/n!$ has zero coefficients, and therefore Px vanishes at all points interior to $(a, R - |a|)$. Within the circle just mentioned take a point b; the point b being a limit-point of the zeros of Px we can use the same reasoning as before and show that Px is zero at every point interior to $(b, R - |b|)$. Thus step by step we can prove that Px vanishes over the whole of its domain (R); and then we see that Px must have zero coefficients (§ 80).

The theorem here established does not exclude the possibility of a series having infinitely many zeros in its domain of convergence. When the zeros are infinite in number the series has zero coefficients, or else the zeros of the series have no limit-point in the domain (R) and at least one limit-point on the boundary of the domain, that is on the circle of convergence itself.

Corollary I. *When two power series in x converge within a circle (R) and are equal for infinitely many values within a circle (R'), where $R' < R$, then the two series are identical; that is, they have the same coefficients.*

This corollary is an immediate consequence of the theorem, and needs no proof.

Corollary II. *The points within the domain of Px at which the series takes a determinate value b_0 have no limit-points interior to the circle of convergence of Px, provided that Px does not reduce to the constant b_0.*

This corollary can be proved by applying the theorem to the series $a_0 - b_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots$, instead of to $\sum\limits_{n=0}^{\infty} a_n x^n$.

In the theorem and in the two corollaries the points which are considered are members of infinite systems confined within

finite regions. When the points of such a system have no limit-points *in the interior of* the associated region, they are said to be *isolated within the region*; the points of a finite system are isolated, whether the region be finite or infinite.

89. Continuation of a Function defined by a Power Series. Let us now return to Taylor's theorem. And first let us show by a simple instance that when from a series Px we deduce a series $P(x-x_0)$ by means of the theorem, the new circle of convergence may cut the old one. Take the geometric series

$$1 + x + x^2 + \ldots,$$

whose sum is $\dfrac{1}{1-x}$ when $|x| < 1$.

Here then $Px = \dfrac{1}{1-x}$, and by differentiation

$$P'x = \frac{1}{(1-x)^2},\quad P''x = \frac{2}{(1-x)^3},\quad \ldots,$$

whence Taylor's theorem takes the form

$$\begin{aligned}\frac{1}{1-x} &= \frac{1}{1-x_0} + \frac{x-x_0}{(1-x_0)^2} + \frac{(x-x_0)^2}{(1-x_0)^3} + \ldots \\ &= \frac{1}{1-x_0}\left[1 + \frac{x-x_0}{1-x_0} + \left(\frac{x-x_0}{1-x_0}\right)^2 + \ldots\right]\end{aligned}$$

The series in parentheses is again a geometric series, and is convergent if $\left|\dfrac{x-x_0}{1-x_0}\right| < 1$, that is, if x is nearer to x_0 than 1 is. Thus the new series defines the function in the circle through 1 whose centre is x_0, while the original series defined the function in the circle through 1 whose centre is 0; and manifestly the new circle will in general cut the old one. Thus ground is gained; we have a new series equal to the old in the region common to their domains, but serving to define the function for an exterior region.

In this very simple case we do not need the aid of power series; we know all about the function $\dfrac{1}{1-x}$ to begin with

The use of the power series will, however, be appreciated as we proceed; the present point is that the domain of the continuation of a series may very well extend beyond the original domain.

Ex. Prove the binomial theorem for $(1-x)^n$, where $|x|<1$ and n is a negative integer, from the series used in this article.

When we have a power series $P(x-x_0)$ with a radius of convergence R_0, and a point x_1 in its domain, we have by Taylor's theorem a new power series in $x-x_1$, say $P(x-x_1)$, which is known to be equal to the old series for all points in the circle $(x_1, R_0-|x_1-x_0|)$. Let the radius of convergence of the new series be R_1; and suppose that, as in the instance just given, the circles of convergence cut each other. We must prove that the two series are equal at all points which are in *both* circles of convergence; these points constitute an open region which can be called the common domain of the two series.

It is sometimes convenient to denote the new series by $P(x-x_0|x_1)$; the notation indicates that the series is deduced

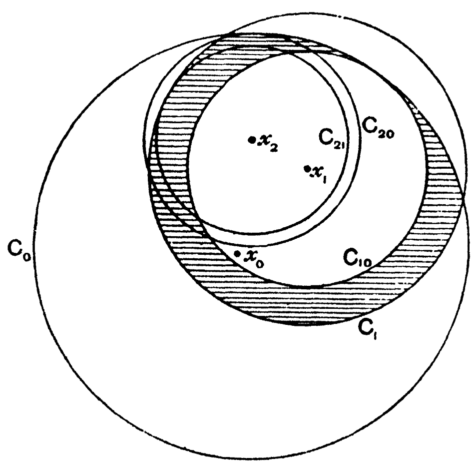


Fig. 41.

(by Taylor's theorem) from the given series $P(x-x_0)$. It is also convenient to denote by C_n the circle of convergence of a

series $P(x-x_n)$, and by C_{nm} the circle which is concentric with C_n and touches C_m internally (fig. 41).

With this notation, we know that our two series are equal within the circle C_{10} and we have to prove that they are equal in the region shaded in the figure; to this shaded region the boundary points on C_0 and C_1 do not belong.

Take x_2 within C_{10} and with it as centre draw a circle C_{20} to touch C_0 internally and a circle C_{21} to touch C_1 internally. We shall need only the smaller of these; let it be C_{21}. We can deduce a series in $x-x_2$ directly from $P(x-x_0)$; let this be $P(x-x_0 \mid x_2)$. Then we know that

$$P(x-x_0) = P(x-x_0 \mid x_2)$$

within C_{20}.

Again we can deduce a series in $x-x_2$ from $P(x-x_0 \mid x_1)$; let this be $P(x-x_0 \mid x_1 \mid x_2)$. Then we know that

$$P(x-x_0 \mid x_1) = P(x-x_0 \mid x_1 \mid x_2)$$

within C_{21}.

Hence within the smaller circle C_{21} both equalities hold. But also we have within C_{10}

$$P(x-x_0) = P(x-x_0 \mid x_1).$$

Therefore in the region common to C_{10} and C_{21} we have

$$P(x-x_0 \mid x_2) = P(x-x_0 \mid x_1 \mid x_2);$$

whence, *as the series are both series in* $x-x_2$, *and* x_2 *is in the region common to* C_{10} *and* C_{21}, they are one and the same series (§ 80).

Hence we have within C_{21}

$$P(x-x_0) = P(x-x_0 \mid x_2),$$

and

$$P(x-x_0 \mid x_1) = P(x-x_0 \mid x_2);$$

whence $P(x-x_0)$ and $P(x-x_0 \mid x_1)$ are equal within C_{21}. The series have been proved equal in that part of the shaded region lying within C_{21}. We now add this part on to C_{10}; and by selecting suitably a new point x_3 in the region so extended, drawing the circles C_{31} and C_{30} and taking the smaller, we again

increase the region of equality of the series; and the process can clearly be continued until the extended region includes all points of the shaded region, that is until the equality at all points of the common domain is established. The question of the equality of the series on the boundary of the domain is left an open one.

There is one possible power series which requires special mention. In the case of $1/(1-x)$, there is an expansion in powers of $x-\infty$ or $1/x$, namely $\sum_{n=1}^{\infty}(-1)^n/x^n$; this series converges at all points exterior to $(0, 1)$; take any element $P(x-x_0)$ of $1/(1-x)$ whose domain lies partly outside $(0, 1)$. At all points outside $(0, 1)$ and in the domain of $P(x-x_0)$ we have $\sum_{n=1}^{\infty}(-1)^n/x^n = P(x-x_0)$; and therefore the series in $1/x$ can be called a continuation of $P(x-x_0)$.

More generally when we have an expansion $P(1/x)$ convergent at all points exterior to (R), and equal to $P(x-x_0)$ at all points in their common domain, then $P(1/x)$ is called a continuation of $P(x-x_0)$.

90. The Analytic Function. An *analytic function* is defined by an aggregate of series composed of a primary series and its continuations; the separate series are called *elements* of the analytic function, and the primary series is called the primary element. Among these *elements* is to be included $P(1/x)$ in case $P(1/x)$ is a continuation of an element $P(x-x_0)$.

Let us see how the value of the analytic function can be found at a point b which lies outside the domain of the primary element.

It is necessary to form a chain of series

$$P(x-a),\ P(x-x_1),\ P(x-x_2),\ \ldots,\ P(x-x_n),$$

in which the first link is the primary element while the last link is a series whose domain includes b. The method of construction of this chain will be understood clearly by examining the way in which the second link is attached to the first.

Let x_1 be situated in the domain (a, R) of $P(x-a)$. Then $P(x-x_1)$ is that series $P(x-a\,|\,x_1)$ which is deduced from $P(x-a)$ by writing $x-a=x_1-a+(x-x_1)$ and using Taylor's theorem. The radius of convergence of $P(x-a\,|\,x_1)$ is at least as great as the shortest distance from x_1 to the circle (a, R). Hence if we take x_1 within the shaded region $(a, R/2)$ we can be certain that a will lie within the domain of the new series in $x-x_1$; so that $P(x-a)$ can be deduced from $P(x-a\,|\,x_1)$ by a direct application of Taylor's theorem.

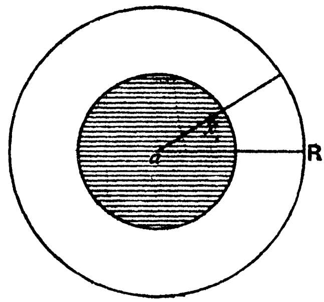


Fig. 42.

Let $P(x-a\,|\,x_1\,|\,x_2)$ be a series in $x-x_2$ which is deduced from

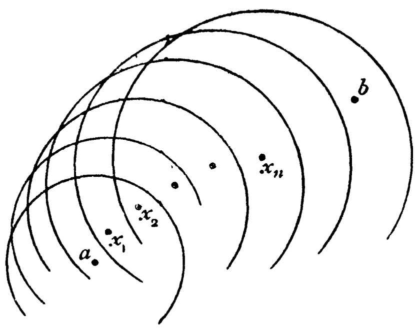


Fig. 43.

$P(x-a\,|\,x_1)$ by the use of Taylor's theorem, and let x_2 be chosen in the region $(x_1, R_1/2)$, where (x_1, R_1) is the domain of $P(x-a\,|\,x_1)$. Continuing in this way it may be possible to reach an element $P(x-a\,|\,x_1\,|\,x_2\,|\,\ldots\,|\,x_n)$ whose domain contains b, and from this final link we can deduce, by immediate continuation, an element $P(x-b)$ whose constant term denotes the value of the analytic function when $x=b$.

In the case of the analytic function $1/(1-x)$ the value at b does not depend on the special chain that happens to be selected; for we can deduce from the series in $x-x_n$ a series in $x-b$ and this series can only be $\sum\limits_{n=0}^{\infty}(x-b)^n/(1-x)^{n+1}$, into which

$x_1, x_2, \ldots, x_n$ do not enter. But with a different analytic function the value at b of the function might very well depend on the manner in which $x_1, x_2, \ldots, x_n$ are interpolated. In this way it is possible to allow for several or even infinitely many values at a point.

The chain described above permitted the passage from the series $P(x-a)$ to the series in $x-b$ by a succession of immediate continuations. The choice of the points $x_1, x_2, \ldots, x_n$ was made with a special view to the reversal of the process. The series in $x-b$ can be chosen as the primary element; for the series in $x-x_n$ is an immediate continuation of it, and the series in $x-x_{n-1}$ is an immediate continuation of the series in $x-x_n$, etc., and finally $P(x-a)$ is an immediate continuation of $P(x-a \mid x_1)$, so that it is possible to pass from the series in $x-b$ to the series in $x-a$ and thence to every element of the analytic function by a chain of immediate continuations. A chain with this property of being available each way may be called a *standard* chain.

We have proved then that *any one of the elements of an analytic function, in particular of the analytic function* $1/(1-x)$, *can be treated as the primary element, all the other elements being deducible from this by the process of continuation.*

This theorem implies that each element contains in germ the whole of the analytic function.

91. General Remarks on Analytic Functions. (1) In the case of the elementary analytic functions the arithmetic expression comes first, the power series second. It is otherwise with a majority of functions. Take for example the case of a differential equation; when it is found impossible to write down in terms of a finite number of known analytic functions a solution of such a differential equation as

$$xy\frac{dy}{dx}=x+y,$$

what is done is this. We select an ordinary point $x=a$ and find y as a series $P(x-a)$, and then deduce the analytic function from $P(x-a)$ by continuations.

(2) The question which is suggested by the consideration of such an equation as

$$\frac{dy}{dx}=f'x,$$

where $f'x$ is a one-valued analytic function, is this: given that $P'x$ is one element of $f'x$, is it safe to infer from the solution $y=Px$ that the equation is satisfied by the analytic function of which Px is merely one element? Or we may ask the closely related question: are the first, second, ..., derivates of the continuations of $P(x-a)$ the same as the continuations of the first, second, ..., derivates of $P(x-a)$? The answer is in each case in the affirmative. Without entering into any details,—for these would carry us far into the subject of differential equations,—we may say that the general principle involved is that a property possessed by one element of an analytic function is possessed also by the others.

(3) It would be a serious misconception of Weierstrass's views to suppose that he wished to employ no other means of investigating the properties of any selected class of functions than those employed in the continued use of Taylor's theorem. In all cases he had recourse to some functional property:—an algebraic equation $f(x,y)=0$, an addition-theorem, a differential equation, and so forth, and in this way gained a control over the subject-matter, that could not have been obtained from the series alone.

(4) The passage from a to b in § 90 was by a succession of steps a to x_1, x_1 to x_2, and so on. The number of interpolated points may be decreased; it is certain that it can be increased as much as we please. There is often an advantage in thinking of a continuous route or *path* from a to b; the meaning then is that the various sets of interpolated points all lie on this line and furnish *standard* chains.

(5) Starting from an initial point a and an initial series $P(x-a)$ and confining the paths to an assigned region it may happen that the value at every point b of the region is unique, whereas this ceases to hold when the region is replaced by a larger one. In this case we have for the first region a one-valued portion of a many-valued analytic function.

92. Preliminary discussion of Singular Points. Analytic functions exhibit every variety of behaviour with regard to continuation beyond the domain of the primary element. The nature of the function depends, in fact, on the nature of the obstacles, called *singular points*, which limit the domains of the series. What these obstacles are will appear further on in more detail; but even at this stage we can give a plausible description of the ordinary obstacles.

In the case of the series $1+x+x^2+\dots$, the obstacle is the point $x=1$, at which $\dfrac{1}{1-x}$ becomes ∞. Naturally, since a power series is defined for each point of its domain, no circle of convergence can enclose this point; but all circles may be expected to pass through it.

In the case of the series $1 + x^2 + x^4 + \ldots, = \frac{1}{1-x^2}$, there are two obstacles, $x = \pm 1$. These and these only are what stop the circles of convergence. A circle of convergence with centre x_0 will pass through *the nearer* of these points; that is, if $x_0 = \xi_0 + i\eta_0$, through 1 when ξ_0 is > 0, through -1 when ξ_0 is < 0, through both when ξ_0 is 0.

In the case of the series for $(1-x)^{1/2}$, namely

$$1 - \frac{1}{2}x + \frac{1}{2^2}\frac{x^2}{2!} - \frac{1 \cdot 3}{2^3}\frac{x^3}{3!} + \ldots,$$

the obstacle is $x = 1$. If the circle of convergence of the series can enclose this point, then when x describes a small circle about 1 (fig. 44) the series is unaltered and the sum is the same, namely $(1-x)^{1/2}$. But $\sqrt{(1-x)}$ is not the same after describing the circle, for the amplitude of $1-x$ has increased continuously through 2π; whence the amplitude of a selected square root has increased continuously through π, and the values $(1-x)^{1/2}$, $-(1-x)^{1/2}$ interchange. Hence the point is not *in* the circle of convergence; it is however *on* it.

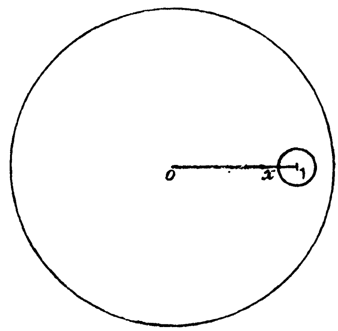


Fig. 44.

In the first case there is no ordinary power series about the point 1, but there is a negative power, namely $-(x-1)^{-1}$.

In the second case there is equally no ordinary power series about the point 1, but there is a power series with a negative power of $x - 1$; since if $x = 1 + x_1$,

$$\frac{1}{1-x^2} = -\frac{1}{2x_1 + x_1^2} = -\frac{1}{2x_1}\left(1 + \frac{x_1}{2}\right)^{-1},$$

or if $|x_1| < 2$,

$$\frac{1}{1-x^2} = -\frac{1}{2x_1} + \frac{1}{4} - \frac{x_1}{8} + \ldots.$$

In the third case there is a fractional power series which has only one term. Thus when about a point x_0 there is no element of an analytic function there may be a series $P\, 1/(x - x_0)$, with a

finite or infinite number of terms, or there may be a Laurent series $P\,1/(x-x_0)+P(x-x_0)$, or again there may be a fractional series such as $P\sqrt[n]{x-x_0}$, or the fractional series may be a Laurent one. This is merely an indication of possibilities, not a classification of them.

In the above cases we know the sum of the given series, that is, we have direct expressions for our functions, and it is to these that we look for information as to the obstacles. This will be observed also in other cases. In general we rely on the theory of this chapter for the regular behaviour of a function, to infer for example its continuity; but for its peculiarities, just those things which mark it off from other functions, we may do well to look in a different direction.

93. Transcendental Integral Functions. When Px has an infinite radius of convergence the series defines a *transcendental integral function.* Its properties resemble in several respects those of the rational integral function, but these two classes of function behave very differently at infinity. There is no need to discuss continuations, for now the single series Px defines by itself the complete analytic function.

Among the simplest transcendental integral functions are those defined by the series $\sum\limits_{n=0}^{\infty} x^n/n!$, $\sum\limits_{n=0}^{\infty}(-1)^n x^{2n+1}/(2n+1)!$, $\sum\limits_{n=0}^{\infty}(-1)^n x^{2n}/2n!$; these functions are discussed in the next chapter.

A special notation Gx is used for integral functions, whether rational or transcendental.

When Gx is not reducible to its first term, there must be points in the finite part of the plane at which $|Gx|$ is greater than any assigned positive number, however large.

Assume, if possible, the existence of a finite upper limit γ for $|Gx|$. Then $A_n \leqq \gamma/\rho^n$, however large ρ may be (§ 78). This requires that $a_1, a_2, a_3, \ldots$ shall all vanish, contrary to supposition. Hence the theorem is true. This important theorem is due to Liouville.

For a *rational integral function* Gx there is a radius R such that for *all* points exterior to (R) we have $|Gx| > \gamma$, where γ is an arbitrarily assigned positive number, however great (§ 56).

94. Natural Boundaries. Another case must be mentioned, where there is a finite radius of convergence, but on every arc of the circle of convergence there is an obstacle. There is then no possible continuation beyond this circle. In this case the circle constitutes what is known as a *natural boundary* for the function. As such functions belong to the higher theory,—where however they present themselves in an unsought way,—we merely give an instance. Lerch (*Acta Mathematica*, vol. x. p. 87) has shown that, if q and r are integers, the point $e^{2q\pi i/r}$ on the circle of convergence $(0, 1)$ of $\sum_{n=1}^{\infty} x^{n!}, = fx$, is an obstacle. For let $x = \rho e^{2q\pi i/r}$ $(\rho < 1)$, and let ρ increase; then as ρ approaches 1 the part of the series from the rth term onwards, namely $\sum_{n=r}^{\infty} \rho^{n!}$, tends to infinity. This would be impossible were the point $e^{2q\pi i/r}$ situated in the interior of any immediate continuation of the power series. It is clear then that there are on the circle $(0, 1)$ infinitely many obstacles, and further that on any arc there are infinitely many obstacles.

Several distinct Analytic Functions defined by one and the same Arithmetic Expression. Tannery has given a very simple example of this peculiarity. Since

$$\lim_{n=\infty} 1/(1 - x^n) = 1, 0,$$

according as $|x| <, > 1$, it follows that the series whose arithmetic expression is

$$\phi x = \frac{1}{1-x} + \left(\frac{1}{1-x^2} - \frac{1}{1-x}\right) + \left(\frac{1}{1-x^4} - \frac{1}{1-x^2}\right) + \dots$$
$$+ \left(\frac{1}{1-x^{2^n}} - \frac{1}{1-x^{2^{n-1}}}\right) + \dots,$$

or

$$\phi x = \frac{1}{1-x} + \sum_{n=1}^{\infty} \frac{x^{2^{n-1}}}{x^{2^n} - 1},$$

has 1 or 0 for its sum under the same circumstances.

Now this series is uniformly convergent in the closed region (R) where R is any positive number < 1; for the remainder

$$r_n x = \frac{-x^{2^n}}{1 - x^{2^n}}$$

is less than $R^{2^n}/(1 - R^{2^n})$ in absolute value. The series therefore fulfils the requirements of Weierstrass's theorem of § 81 and can be transformed into a series Px when $|x| < 1$; similarly when $|x| > 1$ it can be transformed into $P(1/x)$. The arithmetic expression, then, gives parts of two distinct analytic functions within and without the circle. These functions happen to reduce to the constants 1, 0.

Let $f_1 x$, $f_2 x$ be any two one-valued analytic functions of x with a finite number of singular points, then the expression

$$\frac{f_1 x + f_2 x}{2} + \frac{f_1 x - f_2 x}{2} \phi x$$

defines, within and without the circle (0, 1), parts of two distinct analytic functions $f_1 x$, $f_2 x$; the circle itself is not a natural boundary for either of these functions.

CHAPTER XIII.

ANALYTIC THEORY OF THE EXPONENTIAL AND LOGARITHM.

95. The Addition Theorem of the Exponential. The simplest instance of a transcendental integral function is the exponential, defined by

$$y = \exp x = \sum_{n=0}^{\infty} x^n/n!.$$

The test $\mathrm{L}\, A_{n+1}/A_n = 1/R$ (§ 76) shows that here $R = \infty$. One marked peculiarity of this function is that

$$D_x y = \sum_{n=0}^{\infty} x^n/n! = y;$$

it was on this fact that the geometric theory of ch. IV. was based. The function was deduced indeed from this fact, with the condition that $y = 1$ when $x = 0$.

By direct multiplication of power series (§ 83) we have the 'addition theorem':

$$\exp x_1 . \exp x_2 = \exp(x_1 + x_2);$$

for the product of the series for $\exp x_1$ and $\exp x_2$ is

$$1 + (x_1 + x_2) + \left(\frac{x_1^2}{2!} + x_1 x_2 + \frac{x_2^2}{2!}\right) + \left(\frac{x_1^3}{3!} + \frac{x_1^2 x_2}{2!} + \frac{x_1 x_2^2}{2!} + \frac{x_2^3}{3!}\right) + \dots$$

$$= 1 + (x_1 + x_2) + \frac{(x_1 + x_2)^2}{2!} + \frac{(x_1 + x_2)^3}{3!} + \dots\dots\dots$$

In particular if $x_2 = -x_1$, we have

$$\exp x_1 . \exp(-x_1) = \exp 0 = 1,$$

and

$$\exp(-x_1) = 1/\exp x_1.$$

If the argument be an imaginary $i\eta$, and if, separating the series for $\exp i\eta$ into a real and an imaginary series, we write

$$\exp i\eta = \xi' + i\eta',$$

then $$\exp(-i\eta) = \xi \;-i\eta',$$

and by multiplication $$1 = \xi'^2 + \eta'^2.$$

Hence $\exp i\eta$ cannot vanish for a finite η.

That $\exp \xi$, where ξ is real, does not vanish for a finite ξ is clear when $\xi > 0$ from the fact that the series is formed from positive terms beginning with 1; and when $\xi < 0$, $\exp \xi = 1/\exp(-\xi)$, where the denominator is finite.

Since, if $x = \xi + i\eta$, $\exp x = \exp \xi \exp i\eta$ and neither factor is zero, we infer that *the exponential has no zeros.*

When the real number ξ tends to $+\infty$, $\exp \xi$ tends to $+\infty$; hence $\exp -\xi$ tends at the same time to zero. It may seem therefore that we ought to say that $\exp x$ has a zero at infinity, on the same grounds that we say that $1/x$ has a zero at infinity. There is this important difference between the two cases:—As in § 61, we mean by "value of $1/x$ when $x = \infty$" merely "the limit of $1/x$ when $x = \infty$," this limit *being independent of the manner of approach* to the artificial point $x = \infty$; whereas we can assign no such meaning to "value of $\exp x$ when $x = \infty$," for now there is no unique limit (§ 57), that is, no limit independent of the manner of approach. For this reason the value of $\exp x$ at $x = \infty$ is left entirely undetermined. This is a very good illustration of one kind of *singular point* (§ 103).

We are justified in speaking of ∞ as a singular point of e^x. For if there were a series $P(1/x)$ good outside any circle (R) we should have $P(1/\xi) = \exp \xi$ for all values $\xi > R$. But if ξ tend to $+\infty$, $P(1/\xi)$ tends to its constant term whereas $\exp \xi$ tends to $+\infty$, hence the equation cannot be true.

The addition theorem admits of an evident extension, namely

$$\exp x_1 \exp x_2 \ldots \exp x_n = \exp(x_1 + x_2 + \ldots + x_n).$$

In particular if the arguments are all equal to x,

$$(\exp x)^n = \exp nx,$$

where n is a positive integer; and, if $x = 1$,

$$\exp n = (\exp 1)^n,$$

where
$$\exp 1 = 1 + 1/2! + 1/3! + \ldots$$
$$= 2{\cdot}718281828\ldots,$$
the number denoted by e (§ 29).

Let the number of terms x_n be unlimited, but let the series Σx_n be convergent with a sum s; and let $s - s_n = r_n$. We shall prove that
$$\mathrm{L} \exp s_n = \exp \mathrm{L}\, s_n = \exp s.$$

We have
$$s_n = s - r_n,$$
$$\exp s_n = \exp s \,.\, \exp - r_n,$$
$$|\exp s_n - \exp s| = |\exp(-r_n) - 1|\,|\exp s|$$
$$= |-r_n + r_n^2/2! - \ldots|\,|\exp s|$$
$$\leqq (|r_n| + |r_n^2/2!| + \ldots)\,|\exp s|;$$
and
$$|r_n| < \epsilon \text{ when } n > \mu,$$
therefore, for such values of n,
$$|\exp s_n - \exp s| < (\epsilon + \epsilon^2/2! + \ldots)\,|\exp s|,$$
$$< (\epsilon + \epsilon^2 + \ldots)\,|\exp s|,$$
$$< \frac{\epsilon}{1 - \epsilon} \exp s,$$
$$< \epsilon_1,$$
where
$$\epsilon = \frac{\epsilon_1}{\epsilon_1 + \exp s}.$$

Hence when ϵ_1 is arbitrarily selected, ϵ is determined and thereby μ. Hence $\exp s_n$ has a limit, and the limit is $\exp s$. That is, *the addition theorem applies also to a convergent series in the form*
$$L \exp x_1 \,.\, \exp x_2 \,.\, \exp x_3 \ldots = \exp L\,(x_1 + x_2 + x_3 + \ldots).$$

96. The Logarithm. Let
$$1 + z = \exp x.$$

Then
$$z = x + x^2/2! + x^3/3! + \ldots,$$
and $z = 0$ when $x = 0$.

There is therefore one branch of x, which is zero when z is zero, as we knew already (§ 30); the other branches are $2n\pi i$ when z is zero, where n is any integer.

We know by the theory of the reversion of series (§ 84) that that value of x which is zero when $z=0$ can be expressed as $x=P_1 z$.

To determine the coefficients of this series we have

$$D_z x = 1/D_x z = 1/(1+z)$$
$$= 1 - z + z^2 - z^3 + \dots \quad \dots\dots\dots\dots (1),$$

and

$$x = z - z^2/2 + z^3/3 - z^4/4 + \dots \quad \dots\dots\dots\dots (2).$$

Within its circle of convergence $P_1 z$ is equal to this series. It is therefore (§ 80) the same series, and has the unit circle as its circle of convergence.

Now the element (2) is some value of $\log(1+z)$; that is, it is $\operatorname{Log}\rho + i\theta$, where ρ is the absolute value and θ is some amplitude of $1+z$ (§ 30). When z is real, the element is real, and therefore θ is 0. When z moves within the circle of convergence, starting from $z=0$, the element is continuous and therefore is always

$$\operatorname{Log}\rho + i\theta,$$

where θ_0 is the angle turned through by $1+z$. Hence $-\pi/2 < \theta < \pi/2$ (fig. 45). Thus the element is a portion of the chief branch $\operatorname{Log}(1+z)$.

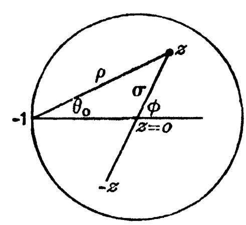


Fig. 45.

Denoting the above restricted value of θ by θ_0, we have

$$\operatorname{Log}\rho + i\theta_0 = z - z^2/2 + z^3/3 - \dots \quad \dots\dots\dots\dots (3),$$

whence, if $|z| = \sigma$ and $\operatorname{am} z = \phi$,

$$\sigma\cos\phi - \tfrac{1}{2}\sigma^2\cos 2\phi + \tfrac{1}{3}\sigma^3\cos 3\phi - \dots = \operatorname{Log}\rho$$
$$= \operatorname{Log}(1+\sigma^2+2\sigma\cos\phi)^{1/2},$$

and

$$\sigma\sin\phi - \tfrac{1}{2}\sigma^2\sin 2\phi + \tfrac{1}{3}\sigma^3\sin 3\phi - \dots = \theta_0$$
$$= \operatorname{Tan}^{-1}\frac{\sigma\sin\phi}{1+\sigma\cos\phi},$$

where $\operatorname{Tan}^{-1}\alpha$ signifies that number in the open interval $(-\pi/2, \pi/2)$ whose tangent is α.

These formulæ hold for all values of ϕ and when $0 \leqq \sigma \leqq 1$, except for the combination $\phi \equiv \pi \pmod{2\pi}$, $\sigma = 1$ (§ 76).

For points on the circle $\phi \equiv 2\theta_0 \pmod{2\pi}$, and the series take the simpler forms

$$\cos\phi - \tfrac{1}{2}\cos 2\phi + \tfrac{1}{3}\cos 3\phi - \ldots = \operatorname{Log}(2\cos\phi_0/2),$$
$$\sin\phi - \tfrac{1}{2}\sin 2\phi + \tfrac{1}{3}\sin 3\phi - \ldots = \phi_0/2;$$

where ϕ_0 is the chief amplitude corresponding to ϕ.

The behaviour of the last series in the neighbourhood of the point π calls for especial notice. As ϕ increases up to π the sum tends to $\pi/2$; as ϕ decreases down to π the sum tends to $-\pi/2$; when ϕ is π the sum is zero. Thus there are distinct limits for the two modes of approach, and the value at the point of discontinuity is the mean of the two.

These real series illustrate another point. Take, for instance,

$$\sin\phi - \tfrac{1}{2}\sin 2\phi + \tfrac{1}{3}\sin 3\phi - \ldots.$$

It is convergent and in spite of discontinuities has a definite sum for every ϕ. But if we differentiate term by term we have the series

$$\cos\phi - \cos 2\phi + \cos 3\phi - \ldots,$$

which is not convergent since $\mathrm{L}\cos n\phi$ is not zero and in fact does not exist. Here we have a good illustration of the possibility of the occurrence of the inequality $D_x \sum_{n=0}^{\infty} f_n x \neq \sum_{n=0}^{\infty} D_x f_n x$; in this case the order of the two limits D_x and $\sum_{n=0}^{\infty}$ is not indifferent. Compare § 86.

We have shown that

$$\operatorname{Log}(1+z) = z - z^2/2 + z^3/3 - \ldots,$$

for points within and on the circle $|z| = 1$, $z = -1$ excepted If we change the sign of z we have

$$\operatorname{Log}(1-z) = -z - z^2/2 - z^3/3 - \ldots$$

for all points within and on the circle except $z = 1$; and the logarithm defined by this series is $\operatorname{Log}\rho' + i\theta_0'$ where $\rho' = |1-z|$ and θ_0' is a value of $\operatorname{am}(1-z)$ such that $-\pi/2 < \theta_0' < \pi/2$.

Now the equation

$$\operatorname{Log} a/b = \operatorname{Log} a - \operatorname{Log} b$$

is true when

$$\operatorname{Am} a/b = \operatorname{Am} a - \operatorname{Am} b,$$

that is, when

$$-\pi < \operatorname{Am} a - \operatorname{Am} b \leqq \pi;$$

in our case $\operatorname{Am}(1+z) - \operatorname{Am}(1-z)$, or $\theta_0 - \theta_0'$, is shown by a

glance at fig. 45 to be $> -\pi/2$ and $< \pi/2$. Hence for all points within and on the circle, the points $z = \pm 1$ excepted,

$$\begin{aligned}\operatorname{Log}\frac{1+z}{1-z} &= \operatorname{Log}(1+z) - \operatorname{Log}(1-z) \\ &= 2(z + z^3/3 + z^5/5 + \ldots) \quad\ldots\ldots\ldots\ldots (4).\end{aligned}$$

If we map the unit circle on to a new plane by writing

$$\frac{1+z}{1-z} = t, \quad z = \frac{t-1}{t+1};$$

and remark that for points on the circle $\operatorname{am} t \equiv \pi/2$ so that t is imaginary, and that $t = 1$ when $z = 0$, we see that the circle maps into the imaginary axis in the t-plane, and the interior of the circle into the half-plane on the right of this axis. Thus for all points t of this half-plane and the imaginary axis, except the points 0 and ∞, we have

$$\operatorname{Log} t = 2\left\{\frac{t-1}{t+1} + \frac{1}{3}\left(\frac{t-1}{t+1}\right)^3 + \frac{1}{5}\left(\frac{t-1}{t+1}\right)^5 + \ldots\right\} \ldots (5).$$

Thus this series for the principal logarithm covers the important case where t is a positive number.

EXAMPLES.

1. For every point considered every term of the series (5) is a power series in t. Hence (§ 87) we can differentiate term by term. Verify that we get in fact

$$D \operatorname{Log} t = 1/t.$$

2. For what values of x is the equation

$$\operatorname{Log} x = \frac{x^2-1}{x^2+1} + \frac{1}{3}\left(\frac{x^2-1}{x^2+1}\right)^3 + \frac{1}{5}\left(\frac{x^2-1}{x^2+1}\right)^5 + \ldots$$

valid?

3. The equation

$$\frac{1}{4}\operatorname{Log} x = \frac{x^{1/2}-1}{x^{1/2}+1} + \frac{1}{3}\left(\frac{x^{1/2}-1}{x^{1/2}+1}\right)^3 + \ldots$$

defines the principal logarithm for all values of x provided we understand by $x^{1/2}$ that square root which lies either in the right-hand half-plane or on the upper half of the imaginary axis.

4. Prove from (4) that when θ is real

$$\sin\theta + \tfrac{1}{3}\sin 3\theta + \tfrac{1}{5}\sin 5\theta + \ldots = \pm\pi/4 \text{ or } 0,$$

distinguishing the cases; and that

$$\cos\theta + \tfrac{1}{3}\cos 3\theta + \tfrac{1}{5}\cos 5\theta + \ldots = \operatorname{Log}\left|\cot\frac{\theta}{2}\right|.$$

5. Prove that

$$1 + 1/3 - 1/5 - 1/7 + 1/9 + 1/11 + \ldots = \pi/2^{3/2},$$
$$1 - 1/3 - 1/5 + 1/7 + 1/9 - 1/11 - \ldots = 2^{1/2}\operatorname{Log}(2^{1/2} + 1).$$

97. The meaning of a^x. In elementary algebra a meaning is assigned, perhaps prematurely, to expressions such as $a^{p/q}$, where p and q are integers, by an appeal to the 'Principle of Permanent Forms.' The explanation leaves matters in this state that the student regards $a^{p/q}$ and $\sqrt[q]{a^p}$ as equivalent symbols, each capable of $|q|$ values. There ensues this difficulty, that $a^{2/2}$ has two values but a^1 has one. When the exponent of a is irrational or complex, the principle referred to becomes quite inadequate.

The fact is that a^x when x is any number follows $\exp x$ in the order of difficulty, though not historically. When we accept $\exp x$, we can proceed without ambiguity as follows.

We saw (§ 95) that when n is a positive integer

$$e^n = \exp n.$$

Whatever number x may be, let

$$e^x = \exp x.$$

Thus $e^{1/2}$ is the positive number $\exp 1/2$; e^π has a single definite value. So we define a^x as $\exp(x \operatorname{Log} a)$, which again is a definite number.

But it must be observed that a different definition is also in use; namely a^x is $\exp(x \log a)$ where $\log a$ is of course any logarithm; here a^x would have in general infinitely many values, for it is $\exp(x \operatorname{Log} a + 2xn\pi i)$; and so would e^x have in general infinitely many values, namely $\exp(x + 2xn\pi i)$.

Observe that the definition we adopt agrees with our previous convention as to $a^{1/n}$, namely (§ 17) that it is not any nth root of a but the chief nth root. For now

$$\begin{aligned} a^{1/n} &= \exp\left(\frac{1}{n} \operatorname{Log} a\right) \\ &= \exp \frac{1}{n}(\operatorname{Log}|a| + i \operatorname{Am} a) \\ &= \exp \frac{1}{n} \operatorname{Log}|a| \,.\, \exp \frac{i}{n} \operatorname{Am} a \\ &= |\sqrt[n]{a}| \operatorname{cis} \frac{1}{n} \operatorname{Am} a. \end{aligned}$$

For a^x we have the series

$$1+x \operatorname{Log} a+(x \operatorname{Log} a)^2/2\,!+\ldots,$$

whence $\lim\limits_{x=0} \dfrac{a^x-1}{x} = \operatorname{Log} a$. Compare § 32.

Ex. 1. Prove that, when a and b are neither 0 nor ∞, the equation $a^x=b$ leads to the equation $b^{1/x}=a$ when x lies in a certain strip of the x-plane which includes the origin.

Ex. 2. Prove that $\lim\limits_{x=0} (1+bx)^{1/x}=\exp b$.

98. The Binomial Theorem.

We have

$$(1-a)^{-x}=\exp -x \operatorname{Log}(1-a),$$

or

$$(1-a)^{-x}=1+u+u^2/2\,!+u^3/3\,!+\ldots,$$

where

$$u=-x \operatorname{Log}(1-a)=xa+xa^2/2+xa^3/3+\ldots,$$

$$|a| \text{ being } < 1.$$

By Weierstrass's theorem (§ 81) this series in u can be regarded as a power series in a, say

$$1+x_1a+x_2a^2/2\,!+x_3a^3/3\,!+\ldots;$$

for each term $u^n/n\,!$ can be regarded as a series in a, uniformly convergent within the circle (0, 1), and the exponential series in u is uniformly convergent in any circle.

We have to determine the coefficient, x_n, of $a^n/n\,!$. It will arise from $u, u^2, \ldots, u^n$, and will consist solely of terms ranging from $(n-1)\,!\,x$ to x^n. That is

$$x_n=x^n+\ldots+(n-1)\,!\,x.$$

But when $-x$ is a positive integer m, where $m<n$, we have the common binomial series for $(1-a)^m$, where the coefficient of $a^n/n\,!$ is 0. Hence x_n has the zeros

$$-1,\,-2,\,\ldots,\,-(n-1),$$

and therefore

$$x_n=x(x+1)\ldots(x+n-1).$$

Hence

$$(1-a)^{-x}=1+xa+\frac{x(x+1)}{1\,.\,2}a^2+\ldots+\frac{x(x+1)\ldots(x+n-1)}{1\,.\,2\ldots n}a^n+\ldots.$$

Thus the binomial theorem is proved *for any exponent, irrational or complex as well as rational*, when $|a|<1$.

But it is only with the definition selected for $(1-a)^{-x}$ in the preceding article that the theorem is generally true; as will be seen by letting $a=0$.

Ex. Apply the test of § 76 to prove that the circle $|a|=1$ is the circle of convergence of the series.

99. The Circular Functions. We proved (§ 32) that when x is real

$$\text{cis}\, x = \exp ix,$$

and for $\exp ix$ we have

$$\exp ix = \Sigma\, (ix)^n/n\,!.$$

Hence equating real and imaginary parts

$$\left.\begin{aligned}\cos x &= 1 - x^2/2\,! + x^4/4\,! - \ldots\\ \sin x &= x - x^3/3\,! + x^5/5\,! - \ldots\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(1).$$

By these same formulæ we define $\cos x$ and $\sin x$ whatever finite number x may be.

Thus $\cos x$ and $\sin x$ are transcendental integral functions, and of course one-valued. Evidently

$$\left.\begin{aligned}2\cos x &= \exp ix + \exp(-ix)\\ 2i\sin x &= \exp ix - \exp(-ix)\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(2).$$

Hence $\sin x$ is zero when

$$e^{ix} = e^{-ix}, \text{ or } e^{2ix} = 1, \text{ or } x = n\pi \text{ or } 0.$$

Thus the zeros are distributed at intervals of π along the real axis, starting from the origin.

So the zeros of $\cos x$ are all on the real axis at intervals of π starting from $\pi/2$.

It is clear that all formulæ of Trigonometry, which can be proved by Euler's formulæ (2), are equally good for complex arguments. For example we prove two addition theorems. We have

$$\begin{aligned}\cos(x_1+x_2) + i\sin(x_1+x_2) &= \exp i(x_1+x_2)\\ &= \exp ix_1 . \exp ix_2\\ &= (\cos x_1 + i\sin x_1)(\cos x_2 + i\sin x_2);\end{aligned}$$

and similarly

$$\cos(x_1+x_2) - i\sin(x_1+x_2) = (\cos x_1 - i\sin x_1)(\cos x_2 - i\sin x_2);$$

whence by addition and subtraction we obtain for complex arguments the familiar formulæ,

$$\cos(x_1+x_2) = \cos x_1 \cos x_2 - \sin x_1 \sin x_2,$$
$$\sin(x_1+x_2) = \sin x_1 \cos x_2 + \cos x_1 \sin x_2.$$

Ex. 1. Prove that

$$\sin^2 x + \cos^2 x = 1,\ \sin(x+\pi/2) = \cos x,\ \cos(x+\pi/2) = -\sin x.$$

Ex. 2. Prove that the square roots of $1-\sin^2 x$ are distinct functions of x. Prove the same for $\sqrt{1 \pm \cos 2x}$.

Ex. 3. Prove generally that the square roots of a transcendental integral function are distinct functions of x when the zeros are all double zeros.

The remaining circular functions are defined as ratios:

$$\sec x = 1/\cos x,\ \csc x = 1/\sin x,\ \tan x = \sin x/\cos x,$$
$$\cot x = \cos x/\sin x.$$

These are *transcendental fractional* functions of x, one-valued by their definitions, which have infinities of the first order at all zeros of their denominators. The secant and tangent are expressed, by division, as power series in x, and are therefore analytic; the cosecant and cotangent are analytic because

$$\csc(x+\pi/2) = \sec x,$$
$$\cot(x+\pi/2) = -\tan x.$$

The hyperbolic functions for real arguments are defined by the equations

$$\cos ix = \cosh x \ = \tfrac{1}{2}(e^x + e^{-x}),$$
$$\sin ix = i \sinh x = \frac{i}{2}(e^x - e^{-x}).$$

Clearly it is not necessary to extend these definitions to complex arguments; the hyperbolic functions are chiefly used to express the real factors of the trigonometric functions of imaginary arguments.

We have as a special case of the addition theorems

$$\begin{aligned}\sin(\xi+i\eta) &= \sin\xi\cos i\eta + \cos\xi\sin i\eta \\ &= \sin\xi\cosh\eta + i\cos\xi\sinh\eta;\\ \cos(\xi+i\eta) &= \cos\xi\cos i\eta - \sin\xi\sin i\eta \\ &= \cos\xi\cosh\eta - i\sin\xi\sinh\eta;\end{aligned}$$

formulæ by which the real and imaginary parts of $\sin x$ and $\cos x$ are set forth.

100. The Inverse Circular Functions. We have solved the equation $\sin x = 0$; that is, we have found the zeros of $\sin x$. To solve the equation $\cos x = \cos a$ we have

$$e^{ix} + e^{-ix} = e^{ia} + e^{-ia};$$

whence $$e^{ix} = e^{\pm ia},$$

and $$ix = 2ni\pi \pm ia,$$

so that $$x = 2n\pi \pm a.$$

To solve the equation $\sin x = \sin a$ we have only to remark that

$$\cos(x + \pi/2) = \cos(a + \pi/2),$$

whence $$x + \pi/2 = 2n\pi \pm (a + \pi/2),$$

and $$x = m\pi + (-)^m a,$$

where m is zero or any integer.

Or we could use the factor formulæ of Trigonometry,

$$\cos x - \cos a = -2 \sin \frac{x+a}{2} \sin \frac{x-a}{2},$$

$$\sin x - \sin a = \quad 2 \cos \frac{x+a}{2} \sin \frac{x-a}{2}.$$

For the equation $\tan x = \tan a$ we have simply

$$\sin(x - a) = 0,$$

whence $$x = n\pi + a.$$

We remark particularly that $\sin x$ takes the same values at the points $a + 2n\pi$, so that $\sin x$ is a *periodic* function, the period being 2π; and so for the cosine. But $\tan x$ has the period π.

Let now $y = \sin x$ and consider how x depends on y. When y is given let there be a value x_0 of x, then all values are given by the formula

$$x = m\pi + (-)^m x_0.$$

But we have to prove that when y is given x can be found; for it might be that $\sin x$ is not capable of all values, just as $\sin \xi$ is not capable of all real values. We know that e^x takes all values since if

$$e^x = y,$$

we have $$x = \operatorname{Log} \rho + i\theta,$$

where $\operatorname{Log} \rho$ and θ can take all real values.

Hence e^{ix} can take all values; and therefore $e^{ix}-e^{-ix}$ can take all values, for if $e^{ix}-e^{-ix}=2b$ then e^{ix} itself is known by solving a quadratic. Thus we arrive at the conclusion that *sin x is capable of all values; that is, if $y = \sin x$, y moves over the whole y-plane.*

But the periodicity of $\sin x$ enables us to consider, instead of the whole x-plane, a *fundamental region* in which one and only one value of x will always be found when y moves over its plane. For whatever x_0 may be assigned there is one and only one of the points $m\pi+(-)^m x_0$ which lies in a vertical strip of breadth π, say in the strip

$$-\pi/2 < \xi \leqq \pi/2;$$

this strip is open on the left but includes its right-hand edge $\xi=\pi/2$. We take this strip as the fundamental region; and calling any point of it x_0 we denote x_0, as depending on y, by

$$x_0 = \operatorname{Sin}^{-1} y,$$

and reserve the notation $x=\sin^{-1}y$ for all the points x; those other than x_0 will similarly be restricted to strips of breadth π, and will when y moves describe paths which are the successive images of the path of x_0 in the lines $\xi=\pi/2$, $3\pi/2$, $5\pi/2$, ..., or $-\pi/2$, $-3\pi/2$, ..., when these lines are regarded as mirrors.

We have then

$$\sin^{-1}y = m\pi + (-)^m \operatorname{Sin}^{-1}y,$$

and we speak of $\operatorname{Sin}^{-1}y$ as *the chief branch* of the function $\sin^{-1}y$. We now find an element of $\sin^{-1}y$ by reverting the series

$$y = x - x^3/3! + x^5/5! - \dots.$$

We have $$D_x y = 1 - x^2/2! + x^4/4! - \dots,$$

$$= \cos x = \sqrt{1-\sin^2 x},$$

the square root being that (perfectly definite) square root which is $\cos x$ and not $-\cos x$. So long as $|\sin x| < 1$ the square root can be expanded by the binomial theorem, and that expansion applies which begins with 1, namely

$$\cos x = 1 - \tfrac{1}{2}\sin^2 x + \dots.$$

Hence we have, when $|y| < 1$,

$$D_x y = \sqrt{1-y^2},$$

where the square root is no longer definite for an unrestricted y, but is definite so long as $|y| < 1$, being that square root which begins with 1 when expanded, i.e. $(1-y^2)^{1/2}$.

Hence
$$\frac{dy}{(1-y^2)^{1/2}} = dx;$$
or
$$dx = dy\left(1 + \tfrac{1}{2}y^2 + \frac{1 \cdot 3}{2 \cdot 4}y^4 + \dots\right),$$
and
$$x = y + \frac{1}{2} \cdot \frac{y^3}{3} + \frac{1 \cdot 3}{2 \cdot 4}\frac{y^5}{5} + \dots,$$
so long as $|y| < 1$.

Since $\mathrm{Sin}^{-1} y$ is 0 with y, this element is $\mathrm{Sin}^{-1} y$ for small enough values of y. We now prove that the element, *throughout its domain*, is $\mathrm{Sin}^{-1} y$. When $\xi = \pm \pi/2$,
$$y = \sin(\pm \pi/2 + i\eta) = \pm \cosh \eta.$$

Thus the edges of the fundamental region map into parts of the real y-axis, from 1 to $|\infty|$ and from -1 to $-|\infty|$. Only when y crosses these parts can x pass out of the fundamental region. Thus the element, which is $\mathrm{Sin}^{-1} y$ when $y = 0$, is $\mathrm{Sin}^{-1} y$ when $|y| < 1$. For other selections of a fundamental region the element would in general belong to two branches. Here the element takes *one* value, namely $-\pi/2$ when $y = -1$, which is not a value of $\mathrm{Sin}^{-1} y$.

Since, when $y = \cos x$, we have
$$y = \sin(x + \pi/2), \quad \text{or} \quad x + \pi/2 = \sin^{-1} y,$$
the function $\cos^{-1} y$ is merely $\sin^{-1} y - \pi/2$ and does not require separate treatment. Nor do $\csc x$ and $\sec x$; for, if $y = \csc x$,
$$\frac{1}{y} = \sin x, \quad \text{and} \quad x = \sin^{-1}\frac{1}{y};$$
while $\sec x = \csc(x + \pi/2)$.

The inverse cotangent depends in the same way on the inverse tangent; but this latter is the simplest of all these functions. For let
$$y = \tan x;$$
then
$$iy = \frac{e^{ix} - e^{-ix}}{e^{ix} + e^{-ix}},$$
$$= \frac{e^{2ix} - 1}{e^{2ix} + 1};$$

whence
$$e^{2ix} = \frac{1+iy}{1-iy},$$
$$2ix = \log\left(\frac{1+iy}{1-iy}\right);$$
whence (§ 96) an element of $\tan^{-1} y$ is
$$x = y - y^3/3 + y^5/5 - \ldots,$$
where $|y| \leqq 1$, the points $\pm i$ excepted.

As a fundamental region we take the strip
$$-\pi/2 < \xi \leqq \pi/2.$$
The points of this strip form the chief branch of $\tan^{-1} y$, this branch is denoted by $\mathrm{Tan}^{-1} y$. When $\xi = \pm \pi/2$,
$$y = \tan(\pm \pi/2 + i\eta) = \pm i \coth \eta.$$
Thus each edge of the strip maps into parts of the imaginary y-axis from i to $i|\infty|$ and from $-i$ to $-i|\infty|$. Hence when $|y| < 1$ x cannot pass out of the fundamental region, and the above element is a part of the chief branch.

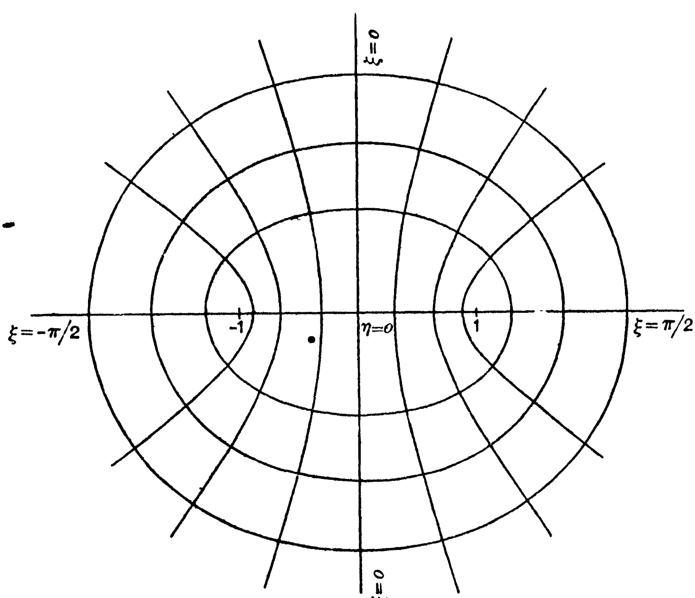


Fig. 46.

101. Mapping with the Circular Functions. We have
$$\sin(\xi + i\eta) = \sin\xi \cosh\eta + i \cos\xi \sinh\eta,$$

whence if $$y = \xi' + i\eta' = \sin x = \sin(\xi + i\eta),$$
$$\xi' = \sin\xi \cosh\eta,\ \eta' = \cos\xi \sinh\eta.$$

If then the x-path is $\xi =$ constant, the y-path is *half* the hyperbola
$$(\xi'/\sin\xi)^2 - (\eta'/\cos\xi)^2 = 1,$$
since $$\cosh^2\eta - \sinh^2\eta = 1, \text{ and } \cosh\eta \geqq 1.$$
It is the right half when $\sin\xi > 0$, and the left when $\sin\xi < 0$.

If the x-path is $\eta =$ constant, the y-path is the ellipse
$$(\xi'/\cosh\eta)^2 + (\eta'/\sinh\eta)^2 = 1.$$
All these conics have the same foci; namely the points $y = \pm 1$.

Therefore they cut at right angles; but this is merely a consequence of the property of isogonality (§ 26).

Since $\cos x = \sin(x + \pi/2)$, the very same curves serve as the map of the orthogonal systems of straight lines when $y = \cos x$.

The corresponding maps for $y = \sec x$ or $\csc x$ are obtained by writing $y' = 1/y$, and are the inverses with respect to the centre of the confocal conics.

For $y = \tan x$ the matter is simpler. We have
$$iy = (e^{2ix} - 1)/(e^{2ix} + 1).$$

Now as x describes a line parallel to an axis, ix describes an orthogonal straight line; and $2ix$ describes a line parallel to this; hence, if $z = e^{2ix}$, the map of the orthogonal lines on the z-plane is (§ 31) a system of rays from the origin and circles round the origin; and since $iy = (z-1)/(z+1)$ the map on the y-plane is a system of coaxial circles, through and about the points which correspond to $z = 0$ and ∞, i.e. $y = \pm i$.

To a vertical x-path, $\xi =$ constant, corresponds a horizontal path for $2ix$, a ray from 0 for z, and an arc between $\pm i$ for y; but to a horizontal x-path from x to $x + \pi$ corresponds a complete circle about the points $\pm i$. Thus for instance to a real x corresponds a real y.

When we make the restriction that y is not to cross a selected arc between $\pm i$, we restrict x to lie in a vertical strip of the x-plane of breadth π; we may regard the y-plane as actually cut along the arc, and if we include one edge of the cut and

exclude the other, then we include one edge of the strip and not the other. On the cut y-plane the function $\tan^{-1}x$ is separated into distinguishable branches. In agreement with the definition of $\mathrm{Tan}^{-1}x$ we make the cut along the imaginary y-axis from i to $i\,|\,\infty\,|$ and from $-i\,|\,\infty\,|$ to $-i$; and we include the right edge of the cut since as any complex x moves from left to right y describes its circle from the left edge of the cut to the right, and we agreed that the right edge of the strip for $\mathrm{Tan}^{-1}x$ was to be counted in.

This notion of a cut, in selecting a branch of a many-valued function, was in common use before the introduction of the Riemann surfaces (ch. XX.); and when these surfaces are employed it is useful in connexion with functions which have thereon *logarithmic* singularities.

The discussion of $\sin^{-1}x$ from this point of view is facilitated by the consideration of a Riemann surface. For

$$x = \sin^{-1} y = \frac{1}{i}\log\left(iy + \sqrt{1-y^2}\right);$$

and here to consider completely the dependence of x on y we first render $\sqrt{1-y^2}$ one-valued for all y's; and then render the logarithm one-valued by a cut.

Ex. 1. Discuss the mapping of parallels to the x-axes by means of $\cot x$.

Ex. 2. Prove that a cut along a complete hyperbola, fig. 46, separates the branches of $\sin^{-1}y$.

Ex. 3. We have, if $y = \cos x$, $2y = z + 1/z$ where $z = e^{ix}$.

Hence when x moves horizontally or vertically, determine the map on the z-plane and thence that on the y-plane.

Ex. 4. From the equation

$$\frac{dy}{\sqrt{1-y^2}} = dx,$$

deduce by equating amplitudes that the tangent of an hyperbola bisects the angle between the focal distances; and by equating absolute values that

$$\int \frac{ds}{\beta} = 2\pi,$$

where the real integral is taken round an ellipse of which s is the arc and β the length of the conjugate radius.

CHAPTER XIV.

SINGULAR POINTS OF ANALYTIC FUNCTIONS.

102. Existence of Singular Points on the Circle of Convergence. The general theorem of which special examples have been given in § 92 is this:—

The circle of convergence (R) *of* Px *must pass through at least one point* x_0 *which has the property that it is impossible to find a power series* $Q(x - x_0)$ *which shall coincide with* Px *in the common domain of the two series.*

The domain of the series cannot contain such a point; what we have to prove is that the points x_0 do not all lie outside the circle of convergence; i.e. at least one lies on that circle. We shall prove first an auxiliary proposition:—

Given that Px *converges in the open region* (R') *and that the radii of convergence of all series* $P(x \mid c)$, *where* c *is any point of this open region, have a lower limit* D, *then the radius of convergence of* Px *is* $R' + D$.

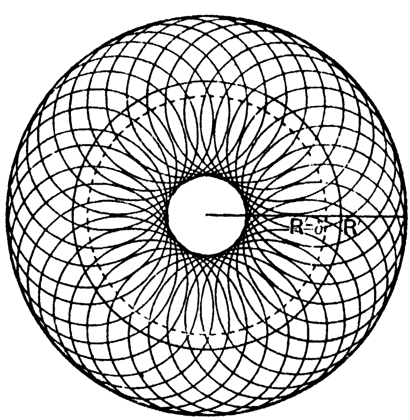


Fig. 47.

The figure shows that every point of a circular region $(R'+D-\delta)$ can be reached by immediate continuations of Px δ being taken as small as we please. We shall use this in the proof of the theorem.

Observe, in the first place, that Px and the aggregate of series $P(x \mid c)$ give one value at each point x in the open region $(R'+D-\delta)$; for any two members of the aggregate coincide in value in their common domain, when the domains happen to overlap. Let us use fx to indicate the common value given by the complete system of series at any point x within $(R'+D)$; it is legitimate to use $R'+D$ instead of $R'+D-\delta$, for x once fixed it becomes possible to take δ so small that x shall lie also within $(R'+D-\delta)$.

We have to show that Px has a domain $(R+D)$, and that

$$fx = Px$$

for all points of this domain. This is done by a double use of Cauchy's inequality (§ 78).

Take c on $(R'-\delta)$ and use D_1, where D_1 is slightly less than D; then if γ be the maximum value of

$$|P(x \mid c)|, = \left| \dots\dots + \frac{P^n c}{n!}(x-c)^n + \dots\dots \right|,$$

on the circle (c, D_1), we have

$$\frac{1}{n!} \mid P^n c \mid \leqq \frac{\gamma}{D_1{}^n}.$$

Now, remembering that $P^n c$ is itself a power series, namely

$$\sum_{m=n}^{\infty} m(m-1)\dots(m-n+1)\, a_m c^{m-n},$$

we get, by a second application of Cauchy's inequality,

$$\frac{m(m-1)\dots(m-n+1)}{n!} A_m \mid c \mid^m \leqq \frac{\gamma \mid c \mid^n}{D_1{}^n}.$$

Now by the binomial theorem,

$$A_m \mid c \mid^m \left(1 + \frac{D_1}{R'}\right)^m = \dots\dots$$

$$+ A_m \mid c \mid^m . \frac{m(m-1)\dots(m-n+1)}{n!} . \left(\frac{D_1}{R'}\right)^n + \dots\dots$$

Hence, writing C for $|c|$,

$$A_m C^m \left(1 + \frac{D_1}{R'}\right)^m \leqq \ldots\ldots + \gamma \left(\frac{C}{D_1}\right)^n \cdot \left(\frac{D_1}{R'}\right)^n + \ldots$$

$$< \sum_{n=0}^{\infty} \gamma \left(\frac{C}{R'}\right)^n, \text{ i.e. } < \gamma \frac{R'}{R' - C}.$$

But $A_m C^m \left(1 + \frac{D_1}{R'}\right)^m$ is the absolute value of the term $a_m x^m$, when x lies on the circle with centre 0 and radius $C\left(1 + \frac{D_1}{R'}\right)$: hence, by the theorem of § 76, the radius of convergence of Px is not less than $C\left(1 + \frac{D_1}{R'}\right)$. But the upper limit of the numbers $C\left(1 + \frac{D_1}{R'}\right)$ is $R'\left(1 + \frac{D}{R'}\right) = R' + D$; hence the radius of convergence is not less than $R' + D$.

It remains for us to prove that the radius of convergence is exactly $R' + D$. Assume if possible that the radius of convergence is $R' + D + D'$ where D' is a number greater than zero. For points c in the open region (R') we can assert that the radii of convergence of the series $P(x|c)$ are all greater than $(R' + D + D') - R'$, for the domain of a series about c extends at least to the circle $(R' + D + D')$. But this is impossible, since it means that the lower limit of these radii is $D + D'$ instead of D. Hence the auxiliary theorem is proved.

There is not much difficulty now in proving the main theorem. Allowing c to range over the interior of $(R + D)$, the radii of convergence of $P(x|c)$ have a lower limit zero. By dividing the circular region $(R + D)$ into small squares (complete and incomplete) as in § 88, we see that for points c in at least one of these compartments the lower limit of the numbers is likewise zero. By increasing indefinitely the number of squares it follows that there is at least one point x_0, within or on $(R' + D)$, such that in every neighbourhood of x_0 the lower limit of the radii of convergence is zero. This point x_0 cannot lie within $(R' + D)$, for the radii in that case are at the worst very nearly equal to $R' + D - X_0$ and therefore cannot have a lower limit zero. Hence the point x_0 lies on $(R' + D)$. Suppose now, if possible,

that there is a series $Q(x-x_0)$ which coincides with Px in the common domain. Then for points c near x_0 we have series $P(x|c)$ which can be deduced from $Q(x-x_0)$ by the use of Taylor's theorem; as the radii of convergence of these series are nearly equal to that of $Q(x-x_0)$ it is impossible that they should have a lower limit zero. Thus the theorem is proved

A point x_0 of the kind considered serves as an obstacle to continuation; we have proved that *on the circle of convergence of any element of an analytic function there is at least one obstacle.* These points are called *singular points or singularities of the function,* and at each of them the function is said to have a *singularity.*

Ex. 1. Prove that 1 is a singular point, and the only one, of the analytic function $1/(1-x)$.

Ex. 2. When Qx/P_0x is expressed as a power series, how far does the domain of this series extend?

103. Non-essential and Essential Singular Points. Let fx be a *one-valued* analytic function. We shall say that it is *analytic about x_0** or *regular at x_0* when it can be represented in the neighbourhood of x_0 by a series $P(x-x_0)$. We shall also say that a function is analytic over a region (closed or open) when it is analytic about each point of that region. The word *holomorphic* is used in the same sense.

Points about which fx is not analytic are therefore *singular points.* For example $1/(x-c)^m$, $e^{1/(x-c)}$ are analytic about all points near $x=c$, but not about $x=c$. More generally the region constituted by points about which fx is analytic is bounded by a finite or infinite number of singular points.

We shall say that the point $x=c$ is a *non-essential singular point* or a *pole* of the analytic function fx when fx can be made analytic about c by multiplication by $(x-c)^m$ where m is a positive integer. When m is the lowest integer which will serve, this integer is called the *order of multiplicity*, or simply the *order* of the non-essential singular point or pole. Interpreting $x-\infty$ as $1/x$ these definitions apply equally to $x=\infty$.

* Adopting, with a slight modification, a suggestion of Professor Bôcher, *Bull. Am. Math. Soc.* vol. iii. p. 89, we use 'analytic about,' in preference to 'regular at.'

Singular points which are not of the kind just described are called *essential* singular points.

Suppose that $x=c$ is a pole of fx, of order m. Then $(x-c)^m fx$ is expressible as $P_0(x-c)$. By division we get

$$fx=\frac{a_{-m}}{(x-c)^m}+\frac{a_{-m+1}}{(x-c)^{m-1}}+\dots$$
$$+\frac{a_{-1}}{x-c}+a_0+a_1(x-c)+a_2(x-c)^2+\dots \quad (1).$$

Since $fx=\infty$ when $x=c$, the pole is an *infinity* of fx, and it is said (in agreement with § 61) to be an infinity of order m. Later (see ch. XVII.) we shall show conversely that an isolated (§ 104) singular point which is an infinity of fx *must* be a pole.

The simple rule that singular points c of the kind (1) are to be called non-essential and *all* others essential is intended to apply only to one-valued functions. If applied to many-valued functions it would make $x=0$ an essential singularity for $\sqrt{x}$, $1/\sqrt{x}$; furthermore we can no longer say that every isolated singularity which is an infinity is necessarily a point of the kind (1); e.g. $x=0$ is an infinity for $1/\sqrt{x}$.

We see from (1) *that we can remove the singularity by subtracting from fx the rational fraction composed of the first m terms.*

When $c=\infty$, we must subtract a polynomial

$$a_{-m}x^m+a_{-m+1}x^{m-1}+\dots+a_{-1}x.$$

Let us consider, in the light of this remark, the ordinary resolution of a rational fraction into partial fractions. Let c be a zero of the denominator, of order m; then c is an infinity of the fraction, of order m, and we have

$$(x-c)^m fx=P_0(x-c);$$

whence by division

$$fx=\frac{a_{-m}}{(x-c)^m}+\dots+\frac{a_{-1}}{x-c}+P_0(x-c)$$

and $fx-\left[\frac{a_{-m}}{(x-c)^m}+\dots+\frac{a_{-1}}{x-c}\right]$ has no infinity at c; we have *removed* the infinity $x=c$. When in this way all infinities, which arise from zeros of the denominator, are removed, we examine $x=\infty$. This is an infinity if the order of the numerator is greater than that of the denominator; to remove it we subtract

a polynomial in x which is determined by division. We have then a rational function without singular points; that is, a constant.

Ex. Express as partial fractions $x^{10}/(x^2-1)^4$ and $x^{10}/(x^2+1)^4$.

104. Character of a One-valued Function determined by its Singularities. Throughout the theory of functions close attention has to be paid to the number and nature of the singular points. It is largely by such a study that the interior structure of a class of functions is best revealed. Let us start with simple cases in which essential singularities are absent.

We know that the singularities of rational fractions are non-essential; is it true conversely that the class of one-valued analytic functions with non-essential singularities only is the same as the class of rational fractions?

This question suggests in turn two questions as to the possible distributions of the non-essential singularities.

(1) Can a one-valued analytic function have no singularities?

(2) Can a one-valued analytic function have infinitely many singularities, all of which are non-essential?

When these two questions have been answered in the negative, the primary question can be disposed of very readily.

The answer to (1) is contained in the following theorem:—

A one-valued analytic function has at least one singular point.

Let us assume, if possible, the existence of a one-valued analytic function fx, not a constant, which has no singularities.

Because fx is analytic about ∞ we have an element

$$a_0 + a_1/x + a_2/x^2 + \dots + a_n/x^n + \dots .$$

Hence, by § 79, $|fx|$ differs very little from A_0 for *all* values of x exterior to some circle (X_0); and therefore there is a circle (R) such that at every point x exterior to (R), we have

$$|fx| < \gamma,$$

where γ is any number greater than A_0.

Now consider the element Px; the radius of convergence must be infinite for otherwise there would be a singular point in

the finite part of the plane, namely on the circle of convergence (§ 102). Because fx is a transcendental integral function there are points exterior to (R) for which

$$|fx| > \gamma.$$

Hence the initial assumption was false, and the first question is answered.

Suppose next that fx has one but only one singular point and that this singular point is non-essential and situated at ∞. There must be an equation

$$fx = a_0 x^m + a_1 x^{m-1} + \ldots + a_{m-1} x + P(1/x),$$

where $a_0 \neq 0$. By subtracting from fx the polynomial

$$a_0 x^m + a_1 x^{m-1} + \ldots + a_{m-1} x$$

we remove the only singularity and get an analytic function which must reduce to a constant, say a_m. Hence

$$fx = a_0 x^m + a_1 x^{m-1} + \ldots + a_{m-1} x + a_m.$$

That is, *a one-valued analytic function fx which has no singularity in the finite part of the plane and a non-essential singularity at infinity is necessarily a rational integral function of x.*

The rational integral function has one singular point and that a non-essential singular point situated at $x = \infty$. The converse theorem shows that one-valued analytic functions with no singular point except $x = \infty$, and that point non-essential, are rational integral functions. Hence *the singularity at ∞,—the only singular point,—of a transcendental integral function is essential.*

We pass on to the consideration of the second question. Given that the singular points,—all non-essential,—are infinite in number, then infinitely many may occur in a circle (R). Suppose that this happens; then the reasoning of § 88 shows that there is some point c of the closed region (R) such that within *every* neighbourhood of c there are infinitely many of these singular points. The impossibility of this will be established by proving the following theorem:—

Given that the one-valued analytic function fx has non-essential or essential singular points other than $x = c$ within every circle

of centre c, no matter how small the radius of this circle, then c itself must be an essential singular point of fx.

The point c cannot be a point about which fx is analytic; for $fx = P(x-c)$ excludes singularities from the immediate neighbourhood of c. Hence the question is narrowed down to the nature of the singularity at c.

Suppose if possible that c is a non-essential singular point, and that

$$fx = (x-c)^{-m} P_0(x-c).$$

Then at every point b near to c, but not coincident with c, $(x-c)^{-m}$ and $P_0(x-c)$ are expressible as power series in $x-b$; hence b is non-singular and there are no singular points in the immediate neighbourhood of b.

As c must be a singular point and as the singularity cannot be non-essential, it follows that c is an essential singular point.

In order then that the infinitely many singular points of fx in the x-plane may be all non-essential, these points must have no limit-points in the finite part of the plane, or as it is often stated they must be *isolated*; for the other supposition requires the existence of essential singularities, contrary to hypothesis.

Before we can give a final answer to the second question, we must first determine whether there may be no essential and yet infinitely many non-essential singular points of which only a finite number lie within any finite region of the plane. The following theorem disposes of this possibility:—

A one-valued analytic function fx with no essential singularities cannot have infinitely many non-essential singularities exterior to (R), where R is arbitrarily great.

The expansion at ∞ is $x^m P_0(1/x)$ where m is zero or a positive integer. The domain of $P_0(1/x$ is the region exterior to some circle, and the only singularity of $x^m P_0(1/x)$ in this domain is at $x=\infty$; calling this circle (R), the theorem is proved.

Having shown then that the number of singularities, supposed all non-essential, is finite, it is very easy to show that fx must be a rational fraction. Let the poles in the finite part

of the plane be $c_1, c_2, \ldots, c_r$, and of orders $m_1, m_2, \ldots, m_r$. Then

$$(x-c_1)^{m_1}(x-c_2)^{m_2}\ldots(x-c_r)^{m_r}fx$$

is a one-valued analytic function which has only one singular point, namely a pole of order m at $x=\infty$. It is therefore equal to a rational integral function of x of degree m, and the required result follows by division.

Attention is called to the fact that the character of a class of functions has been determined from a knowledge of two of their properties, namely that they are one-valued and that their singularities are non-essential.

Reviewing what has been said as to singular points and drawing some evident conclusions, we see that

(1) At an ordinary point $x=c$, $\lim\limits_{x=c}(x-c)fx=0$.

(2) At a pole $x=c$ of order m, $\lim\limits_{x=c}(x-c)^m fx$ exists and is neither 0 nor ∞.

(3) The point c is an isolated singular point for each of the functions $\cos(x-c)+e^{1/(x-c)}$, $\sin(x-c)+\sin 1/(x-c)$; and for each of these functions we have an expansion $\sum\limits_{n=-\infty}^{\infty} a_n(x-c)^n$. Here it is impossible to get rid of the negative terms by multiplying by $(x-c)^m$. More generally let fx be a one-valued analytic function with an isolated singular point,—supposed essential,—at c; and let fx be expressible for all points x near enough to c by

$$fx=G\left(\frac{1}{x-c}\right)+P(x-c),$$

then it is impossible to destroy the negative powers by multiplication by $(x-c)^m$. Later on (in ch. XVII.) we shall see that fx *must* be expressible in the above form under the given suppositions.

(4) When $x=c$ is a pole of fx of order m it is an ordinary point of $1/fx$. This suggests the question whether when $x=c$ is an essential singular point of fx, it is also an essential singular point of $1/fx$ or not. The question must be

answered in the affirmative; for let $x=c$ be (i) an ordinary point or (ii) a pole of $1/fx$, then it is (i) an ordinary point or a pole, or (ii) an ordinary point of fx, contrary to hypothesis.

(5) Strictly speaking there is no "value" for fx at a singular point $x=c$. But if the singularity be non-essential, $\lim\limits_{x=c} fx$ exists and is unique for every mode of approach of x to c, namely it is the number ∞. To indicate this it is usual to say that the value of fx at a non-essential singular point is ∞.

It can be proved that $e^{1/x}$ takes every value except 0 and ∞ near $x=0$, that csc $1/x$ takes every value except 0 near $x=0$, and that there are other one-valued functions which take every value near $x=0$. Picard has proved that in general a one-valued analytic function fx takes, near an essential singularity which is not a limit-point of other essential singularities, any assigned value an infinity of times; at most there are two exceptional values. The proof cannot be given here; but in § 123 we shall prove a cognate theorem.

105. Transcendental Fractional Functions. Having proved that one-valued analytic functions whose singularities are non-essential are rational integral functions, we must introduce essential singularities if we are to get new functions. The simplest case has been considered already, namely one singularity and that essential and situated at ∞. The function is a transcendental integral function Gx. When the singular point is at c instead of at ∞ we must replace Gx by $G(1/(x-c))$.

Let G_1x, G_2x be two integral functions, one but not both of which may be rational. The quotient G_1x/G_2x may be rational as in the case of xe^x/e^x; or it may be transcendental and integral, as in x/e^x. In other cases it is called a *transcendental fractional function* of x. An example is tan x.

The only singularities of a transcendental fractional function in the finite part of the plane are non-essential ones due to the zeros of the denominator; and there is an essential singularity at ∞. Conversely

Every one-valued analytic function fx which has only non-essential

singularities in the finite part of the x-plane and has $x=\infty$ for an essential singularity, is a transcendental fractional function of x.

(1) Let fx have a *finite* number of non-essential singularities in the finite part of the plane, say the poles $c_1, c_2, \ldots, c_r$ of orders $m_1, m_2, \ldots, m_r$; then

$$fx = \frac{Gx}{(x-c_1)^{m_1}(x-c_2)^{m_2}\ldots(x-c_r)^{m_r}}.$$

(2) Let fx have infinitely many non-essential singularities; then these points must have no limit-points or in other words must be *isolated* in the finite part of the plane, otherwise there would be essential singularities other than $x=\infty$.

(3) Let fx have infinitely many non-essential singularities, each of order 1. These can be named $c_1, c_2, c_3, \ldots, c_n, \ldots$, where no c exceeds any subsequent c in absolute value and $\mathrm{L}\, c_n = \infty$; for within any circle (R) the number of singular points is finite. *Assuming that there exists a function G_2x whose zeros are simple and situated at the c's*, then $fx \times G_2x$ has at most one singular point and that situated at infinity, so that

$$fx \times G_2x = G_1x,$$

and

$$fx = G_1x / G_2x.$$

Here then we are confronted with the problem of the construction of G_2x; we shall consider this in the next chapter, and thereby complete the proof of the theorem.

106. Limit-points of Zeros. We have seen that a limit-point of non-essential singularities of a one-valued analytic function fx is an essential singularity of the function. We now show that an analogous theorem holds for the zeros of fx. Suppose that c is a limit-point of the zeros of fx, and let us consider whether c can be either a point about which fx is analytic or a non-essential singular point of fx. In the former case we have

$$fx = P(x-c)$$

in the neighbourhood of c and therefore, by § 80, fx reduces to a constant. Excluding this case, we see that $x=c$ must be

a singular point. Assume, if possible, that it is a non-essential singular point, and consider the function $1/fx$ The zeros of fx are infinities of $1/fx$, also c itself is non-singular for $1/fx$. Hence $1/fx$ has infinitely many non-essential singularities at points that have $x=c$ as a limit-point, without at the same time having an essential singular point at $x=c$. This is impossible by § 104.

We draw from what precedes the important inference that the zeros of a transcendental integral function Gx have no limit-points in the finite part of the plane:—in other words are isolated in the finite part of the plane.

107. Deformation of Paths. Let us understand by a *circuit* in the x-plane any closed path which does not intersect itself. A circuit divides the entire plane into two regions,—an inner and an outer; each of these is said to be *simply connected.* The surfaces of (1) a sheet of paper without holes, (2) one with holes, provide examples of surfaces which are respectively simply and multiply connected. We shall have other examples later (§ 152).

Suppose now that a region Γ in the x-plane is simply connected and that $P(x-x_0)$ is any power series about some assigned point x_0 of Γ. From this power series we can construct for the region Γ a portion of an analytic function fx, which we shall term a *localized function* ϕx, by means of continuations along all possible chains that lead from the point x_0 to the various points of Γ. The aggregate of resulting power series constitutes the function ϕx. It is often very necessary to know whether ϕx is or is not one-valued in Γ Obviously it is one-valued if fx is one-valued; but fx may be many-valued and yet ϕx one-valued. For example fx may be $\log x$, Γ a circular region which does not contain the origin.

Guided by this latter example let us take a region Γ in which the component series of the aggregate ϕx meet with no obstacles to ϕx due to singular points of fx. We can easily secure that this shall be the case by imposing the condition that every path that starts from x_0 and lies wholly in Γ shall furnish a

standard chain from x_0 to the other extremity such that the corresponding radii of convergence shall all be greater than δ (>0). This excludes the possibility of there being any obstacles to ϕx in Γ; such an obstacle would mean $\delta = 0$.

We wish to prove that ϕx is one-valued in Γ, and this will be done by showing that when x describes any two paths C_1 (or *ahb*) and C_2 (or *akb*) which lie wholly within Γ, then ϕx starting from an element at a will arrive by either route at the same final value at b. This theorem may be enunciated in the following manner:—*The difference between the final and initial values of ϕx, or the change in ϕx, when x describes any circuit ahbka in Γ, is zero.*

In proving this theorem we shall use paths instead of chains in order to simplify the exposition as much as possible (§ 91). The proof depends on the obvious propositions:—

(1) The change of a power series when x describes a circuit in the domain of that series is zero.

(2) When a path is described successively in opposite directions, for example *ahb* and then immediately afterwards *bha*, the total change of ϕx for the compound path is zero.

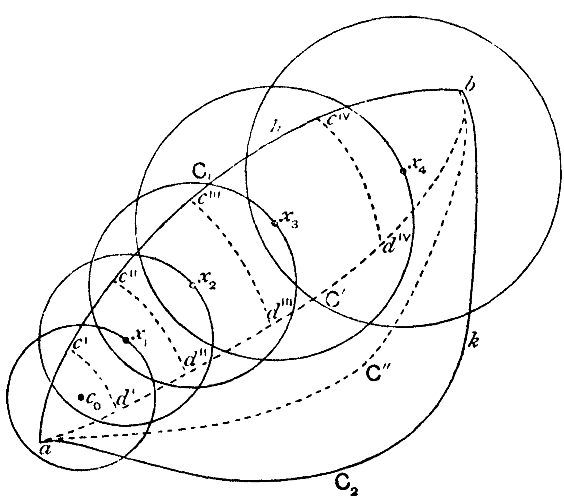


Fig. 48.

On C_1 interpolate points $c', c'', \ldots, c^{(n)}, c^{(n+1)}$, as in fig. 48, where $n=3$; and connect a, b by a path C' which also lies

within these circles and is intermediate in position between C_1 and C_2. On this new path interpolate points d', d'', ..., $d^{(n)}$, $d^{(n+1)}$, as in the figure. The change of ϕx when x goes along C' from a to b is equal to the sum of the changes from a to d', d' to d'', d'' to d''', etc., and these in turn can be replaced by the changes along $ac'd'$, $d'c'c''d''$, $d''c''c'''d'''$, etc., since the description by ϕx of circuits $ac'd'a$, $d'c'c''d''d'$, $d''c''c'''d'''d''$, ..., each of which lies wholly in the domain of one element, must restore the initial value. The paths $c'd'$, $c''d''$, ... are described twice in opposite directions; and therefore the corresponding changes destroy one another. The remaining parts make up C_1. Hence the change of ϕx along C_1 is equal to the change along C'. The change along C' is equal to the change along a path C'' still nearer to C_2, and so on. By means of a *finite* number of intermediate paths C_1 can be converted into C_2, for the original hypotheses imply that all the domains associated with ϕx have radii with a lower limit greater than zero. Hence the theorem is proved.

The theorem of this article shows that when an analytic function starts at a with a given value and passes to b along two routes C_1, C_2, the final values will be equal whenever C_1 can be deformed continuously into C_2 without meeting a singular point. But when fx is one-valued the theorem that ϕx is one-valued is self-evident; also the restriction prohibiting a passage over singular points is unnecessary. The essence, then, of the theorem lies in its application to many-valued functions; from these many-valued functions fx are extracted localized one-valued functions ϕx, and within the regions attached to the functions ϕx the change in passing from a to b is independent of the path.

When a circuit C contains one singular point c the change of fx round the circuit may or may not be zero. It is usual to contract C in this case into a small circle about c, and connect the change round C with that round the circle. Fig. 49 shows how this is done. The sum of the changes over efg, gh, hkl, le is zero. Let el, gh coincide; then the changes along gh, le

need not necessarily cancel each other, for these two lines are not described consecutively. In the case of $\log(x-c)$ they do cancel, despite the circumstance that the values of $\log(x-c)$ at l, h differ by $2\pi i$. Whenever this happens the change over C is equal to the change over the small circle round c.

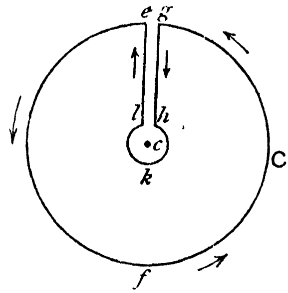


Fig. 49.

The case where C contains several singular points is treated in a similar way.

108. The Logarithm of Gx. If c is a zero of Gx, then we have $Gx = P_m(x-c)$, m being the *order* of the zero. If we remove the factor $(x-c)^m$ the other factor is a series convergent for every x, and itself defines a transcendental integral function G_1x Let

$$P_m(x-c) = a_m(x-c)^m + a_{m+1}(x-c)^{m+1} + \dots,$$

then $$G_1x = a_m + a_{m+1}(x-c) + \dots.$$

Hence $$G_1x = a_m[1 + P_1(x-c)],$$

and, for points so near to c that $|P_1(x-c)| < 1$,

$$\text{Log}[1 + P_1(x-c)] = P_1(x-c) - \tfrac{1}{2}[P_1(x-c)]^2 + \tfrac{1}{3}[P_1(x-c)]^3 - \dots$$
$$= \text{a power series in } x-c.$$

Therefore $\log G_1x = \log a_m + P(x-c)$,

and $$\log Gx = m\log(x-c) + \log a_m + P(x-c).$$

Hence if x describes a circle about c*, so small as to enclose no other zero, $\log Gx$ increases by $2m\pi i$.

When c is a zero of Gx we see that it is a singular point of $\log Gx$; on the other hand when c is not a zero it is not a singular point of $\log Gx$. Hence can be inferred the behaviour

* When we speak of describing a circle about a point it is to be understood that the circle is described positively, unless the contrary is expressly stated. The plane of numbers divides space into two parts; calling one of these the upper part, let an observer in the upper part make a circuit on the plane. The circuit divides the plane into two regions. The region on his *left* is described *positively*; that on his *right*, *negatively*. And the circuit is said to be positive with regard to any point in the left-hand region, and negative with regard to any point in the right-hand region.

of $\log Gx$ for any circuit in the x-plane. For the circuit can be deformed into small circles round each zero of Gx within the circuit. Hence *the change of* $\log Gx$ *is* $2\pi i$. (*the sum of the orders of the enclosed zeros*).

Conversely if there is no change of $\log Gx$ when x makes a circuit, there are no zeros in the circuit.

For example the polynomial

$$Gx = x^m + a_1 x^{m-1} + \ldots + a_m$$

is for large enough values of $|x|$ of the form

$$x^m (1 + P1/x)$$

where $|P1/x| < 1$. Hence $\log Gx$ increases by $2m\pi i$ when x describes a large enough circle, and we infer that the polynomial has m zeros whose total order is m; there may be m simple zeros, one zero of order m, or any intermediate case.

This is an independent proof of the fundamental theorem of algebra (§ 59).

When a function Gx has no zeros, any selected logarithm has no finite singular points and is therefore itself an integral function G_1x (rational or transcendental). Hence *the general expression for a function with no zeros and no singular point except* ∞ *is* e^{G_1x}. Here G_1x is *any* integral function, rational or transcendental. The special case of e^x has been already mentioned (§ 95).

The argument of this article applies much more generally. If a one-valued function has an isolated infinity c, so that near c it is expressed by

$$fx = (x-c)^{-m} P(x-c),$$

then as before $\log fx = -m \log(x-c) + P(x-c)$,

and a small circle about c increases $\log fx$ by $-2m\pi i$. If a circuit encloses zeros and infinities but no other singular points, it can be deformed into small circles round these zeros and infinities and *the change of* $\log fx$ *when* x *describes the circuit is* $2\pi i$. (*total order of the zeros* – *total order of the infinities*).

CHAPTER XV.

WEIERSTRASS'S FACTOR-THEOREM.

109. Infinite Products. Let $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ be a sequence of positive numbers, less than unity. Then

$$(1-\alpha_1)(1-\alpha_2) > 1-\alpha_1-\alpha_2,$$

$$(1-\alpha_1)(1-\alpha_2)(1-\alpha_3) > 1-\alpha_1-\alpha_2-\alpha_3,$$

and so on.

Hence if the series $\Sigma\alpha_n$ has a sum s, the products $\prod\limits_1^n(1-\alpha_m)$ form a sequence of numbers which (1) do not increase, (2) remain greater than $1-s$. Hence they have a limit; and the infinite operation $\Pi(1-\alpha_n)$ is convergent; the limit is called the product and is itself often denoted by $\Pi(1-\alpha_n)$. Thus

The product $\Pi(1-\alpha_n)$ *exists when* $\Sigma\alpha_n$ *is convergent.*

When positive numbers $\beta_1, \beta_2, \ldots, \beta_n, \ldots$, less than unity, converge to β, where $\beta > 0$, then their real logarithms converge to $\operatorname{Log}\beta$.

For this is implied in the continuity of $\operatorname{Log}\beta$, and the continuity follows from the expression as a power series

$$\operatorname{Log}\beta = \beta - 1 - \tfrac{1}{2}(\beta-1)^2 + \ldots.$$

Hence, in our case, $\operatorname{Log}\Pi(1-\alpha_n)$ has a limit, that is, $\Sigma\operatorname{Log}(1-\alpha_n)$ has a limit, on the supposition that $\Sigma\alpha_n$ has a limit.

Ex. Prove $\mathrm{L}\operatorname{Log}\beta_n = \operatorname{Log}\mathrm{L}\beta_n$ from the formula

$$\operatorname{Log}|a/b| = \operatorname{Log}|a| - \operatorname{Log}|b|.$$

Next for any numbers a_n, where $|a_n| = \alpha_n < 1$, we have

$$1 + a_n = \exp \operatorname{Log}(1 + a_n),$$

where the logarithm is defined by the series

$$a_n - a_n^2/2 + a_n^3/3 - \ldots.$$

If $\Sigma \operatorname{Log}(1 + a_n)$ is convergent, say to s, then the addition theorem of the exponential (§ 95) holds, and

$$\mathrm{L} \prod_1^n (1 + a_m) = \exp s.$$

Now $\Sigma \operatorname{Log}(1 + a_n)$ can be written as a double series, which, if we replace each term by its absolute value, becomes

$$\begin{aligned} &\alpha_1 + \alpha_1^2/2 + \alpha_1^3/3 + \ldots \\ &+ \alpha_2 + \alpha_2^2/2 + \alpha_2^3/3 + \ldots \\ &+ \alpha_3 + \alpha_3^2/2 + \alpha_3^3/3 + \ldots \\ &+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \end{aligned}$$

the nth row converging to $-\operatorname{Log}(1 - \alpha_n)$.

Now let the first column $\Sigma\alpha_n$ converge. Then, as we have proved, $\Sigma \operatorname{Log}(1 - \alpha_n)$ converges. Thus in our array of positive terms the sums of the rows converge. Hence:

Σ Log$(1 + a_n)$ is absolutely convergent when $\Sigma |a_n|$ is convergent.

In this case we say that the product $\Pi(1 + a_n)$ whose factors can now be arranged in any order, is *unconditionally convergent.*

For example, $\Pi(1 - x^2/n^2)$ is unconditionally convergent, because $\Sigma\, 1/n^2$ is convergent.

But if $\Sigma\alpha_n$ is not convergent while $\Sigma\alpha_n^2$ is convergent, then the above proof applies to $\Sigma[\operatorname{Log}(1 + a_n) - a_n]$. For example let $a_n = -1/(n+1)$.

Then $$\Sigma\left(\frac{1}{n+1} + \operatorname{Log}\frac{n}{n+1}\right) \text{ is convergent;}$$

that is, $\mathrm{L}[1/2 + 1/3 + \ldots + 1/(n+1) - \operatorname{Log}(n+1)]$ is a finite number. Hence, adding 1 and changing n to $n-1$, we have Euler's result:

$$\mathrm{L}[1 + 1/2 + \ldots + 1/n - \operatorname{Log} n] \text{ is a finite number, } \gamma.$$

For the calculation of this constant γ, which is $\cdot 577215665\ldots$, we may refer to Adams's *Works*, vol. i. p. 459.

110. Construction of Functions Gx **with assigned Zeros.** The well-known theorem that a polynomial of degree n can be expressed in the form

$$C(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n),$$

suggests naturally these two questions:—

(1) How far is a function Gx determined by its system of zeros?

(2) Given a system of numbers $a_1, a_2, a_3, \ldots, a_n, \ldots$ arranged in ascending or stationary order of absolute values and such that $\mathrm{L}\, a_n = \infty$, is it possible to construct a transcendental integral function Gx which shall be zero of order one at the a's and have no other zeros?

NOTE. The zeros are assumed to be of order one; zeros of higher multiplicity are accounted for by the usual device of making several a's coincide.

I. **Functions** Gx **with no Zeros.** We have proved (§ 108) that *a function Gx with no zeros is expressible in the form e^{G_1x}, where G_1x is an integral function, rational or transcendental;* and accordingly we have the following theorem which contains a complete answer to the first question:—

The number of transcendental integral functions with the same system of zeros as an assigned transcendental integral function Gx is unlimited, but these functions are all of the form $Gx\,e^{G_1x}$, where G_1x is a rational or transcendental integral function.

II. **Functions** Gx **with a Finite Number of Zeros.** It follows from this theorem that the general form for a transcendental integral function with a *finite* number n of zeros, all simple, is

$$(x-a_1)(x-a_2)\ldots(x-a_n)\,e^{Gx}.$$

III. **Functions** Gx **with infinitely many assigned Zeros.** Let us now consider the second question. The answer is furnished by the following theorem of Weierstrass's:—

There exists a transcendental integral function Gx whose zeros are the points a_n and are simple.

We proceed to the discussion of this theorem.

111. Weierstrass's Primary Factors. $\prod_{n=1}^{\infty}(1-x/a_n)$, which appears to satisfy the requirements, may diverge. Weierstrass replaced the factor $1-x/a_n$ by what he called *a primary factor* $(1-x/a_n)\,e^{\phi_n(x/a_n)}$, where $\phi_n x$ denotes a polynomial in x. The primary factor is to be treated as a whole, and like $1-x/a_n$ it has one zero and one singular point (at ∞).

We shall now show how to choose $\phi_n x$. Because $x=e^{\operatorname{Log} x}$, the typical term, say E_n, of the product

$$\prod_{n=1}^{\infty}(1-x/a_n)\,e^{\phi_n(x/a_n)}$$

can be written $$e^{\operatorname{Log}(1-x/a_n)+\phi_n(x/a_n)},$$

and this can be transformed in turn into

$$\exp-\left\{\frac{1}{m_n+1}\left(\frac{x}{a_n}\right)^{m_n+1}+\frac{1}{m_n+2}\left(\frac{x}{a_n}\right)^{m_n+2}+\dots\right\},$$

if $\phi_n x$ be $\sum_{n=1}^{m_n} x^n/n$. Let X, A_n denote $|x|$, $|a_n|$; then to secure convergence we must have $X<A_n$. The series $-\{\dots\}$ will be called the truncated series for $\operatorname{Log}(1-x/a_n)$.

Suppose that a_r is the last of the set $a_1, a_2, \dots$ that lies in any assigned closed region (R); then for all points of this region we can write

$E_{r+1}\,E_{r+2}\,E_{r+3}\dots$ to infinity

$$=\exp\sum_{n=r+1}^{\infty}[\text{the truncated series for } \operatorname{Log}(1-x/a_n)]$$

$$=\exp-\left(\sum_{n=r+1}^{\infty}\sum_{p=m_n+1}^{\infty}x^p/pa_n^p\right).$$

Our first concern is with the double series

$$\sum_{n=r+1}^{\infty}\sum_{p=m_n+1}^{\infty}x^p/pa_n^p \quad\dots\dots\dots\dots(1),$$

for which we suppose $X\leqq R$; the convergence of the product of the E's depends on the convergence of the series. We shall prove that this double series can be converted (by summation by

columns) into a power series in x whose radius of convergence is not less than R; it will then follow that exp (– the double series) has no zero in the closed region (R).

The following lemma will be needed in the discussion of the double series (1).

LEMMA. *Given an infinite sequence of unequal numbers* $a_1, a_2, a_3, \ldots, a_n, \ldots$, *arranged according to increasing or stationary absolute values and such that* $\mathrm{L}\, a_n = \infty$ *and* $|a_1| > 0$, *then it is always possible to associate with each number* a_n *of the system a positive integer* m_n *which will make*

$$\sum_{n=1}^{\infty} \left| \frac{1}{a_n}\left(\frac{x}{a_n}\right)^{m_n} \right| \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

converge for every finite value of x.

The values $m_n = n - 1$ satisfy the requirements of the lemma, for they make the ratio of the $(n+1)$th to the nth term tend to the limit 0, and therefore make the series (2) converge. It is evident that many other choices could be made for the numbers m_n. We shall now use this lemma.

Let us take $m_n = n - 1$. The sum of the series of absolute values in the row containing x/a_n is increased by the suppression of the coefficients $1/p$. The resulting series of absolute values $\sum\limits_{p=n}^{\infty} (X/A_n)^p$ is geometric and has for its sum

$$\left(1 - \frac{X}{A_n}\right)^{-1}\left(\frac{X}{A_n}\right)^n, \quad < \left\{1 - \frac{R}{A_{r+1}}\right\}^{-1}\left(\frac{R}{A_n}\right)^n$$

Hence the row in X/A_n is convergent and has a sum less than $A\,\dfrac{R^{n-1}}{A_n{}^n}$, where $A, = R\left\{1 - \dfrac{R}{A_{r+1}}\right\}^{-1}$, is independent of n. But the series of numbers $\dfrac{R^{n-1}}{A_n{}^n}$ is convergent by the lemma; hence the double series (1) with X, A_n for x, a_n is convergent.

Adding this double series by columns we get a power series in x. Let R_1 be a number slightly less than R; then within the closed region (R_1) there are not more than r a's, say there are q a's. The first q factors of the product $\prod\limits_{n=1}^{\infty} E_n$, when multiplied

together, give a transcendental integral function, and the remaining factors give a power series whose radius of convergence is greater than R_1. Hence the complete infinite product is equal to a power series which converges uniformly in the closed region (R_1). By making R increase indefinitely, we make R_1 also increase indefinitely. Hence the infinite product can be expressed as a power series with an infinite radius of convergence, that is, it is equal to a transcendental integral function. Furthermore since $E_{r+1}\ E_{r+2}\ E_{r+3}\ \ldots$ to infinity $=\exp$ (a power series with a radius of convergence not less than R), $E_{r+1}\ E_{r+2}\ E_{r+3}\ldots$ cannot have zeros within (R). The zeros of $\prod\limits_{n=1}^{\infty} E_n$ within (R) are therefore $a_1, a_2, \ldots, a_r$.

We see then that *the most general function* Gx *whose zeros occur at the assigned points,* $a_1, a_2, a_3, \ldots, a_n, \ldots,$ *defined as above, is necessarily of the form*

$$e^{Gx}\prod_{n=1}^{\infty}\left(1-\frac{x}{a_n}\right).e^{\frac{x}{a_n}+\frac{1}{2}\left(\frac{x}{a_n}\right)^2+\ldots+\frac{1}{n-1}\left(\frac{x}{a_n}\right)^{n-1}}$$

This is the complete solution to the question proposed in § 105. For special distributions of zeros it is possible that the requirements of the lemma may be satisfied by giving to $m_1, m_2, \ldots, m_n, \ldots$ a common constant value. Naturally this value is taken as small as possible. Let the integer k be such a value; then $\Sigma\, 1/A_n{}^k$ is to converge, whereas $\Sigma\, 1/A_n{}^{k-1}$ is to diverge. The double series in this case arises from the array

$$\begin{array}{lll} x/a_1 & x^2/2a_1^2 & x^3/3a_1^3 \;\ldots\ldots \\ x/a_2 & x^2/2a_2^2 & x^3/3a_2^3 \;\ldots\ldots \\ x/a_3 & x^2/2a_3^2 & x^3/3a_3^3 \;\ldots\ldots \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \end{array}$$

by suppressing the first $k-1$ columns which give divergent series and then the first r rows.

When k is available Gx will be said to be of the *grade* $k-1$; the usual term *class* is open to objections.

112. The Factor-formula for sin πx. In illustration of Weierstrass's factor-theorem let us suppose that the a's are

$\pm 1, \pm 2, \ldots$ and that 0 is a simple zero. Then $\Sigma (A_n)^{-k}$ converges when $k=2$ and diverges when $k=1$. The associated functions are therefore of the first grade.

The most general transcendental integral function with zeros at these points is

$$e^{Gx} x \cdot \prod_{n=-\infty}^{\infty}{}' (1 - x/n) e^{x/n},$$

where the accent denotes here as elsewhere that a special value of n (in this case $n=0$) is omitted because it gives an infinite factor or term.

The formula can be simplified by combining the primary factors which arise from n and $-n$; it then becomes

$$e^{Gx} x \prod_{n=1}^{\infty} (1 - x^2/n^2).$$

The simplest function of this nature is when e^{Gx} is a constant.

We know that $\sin \pi x$ is a transcendental integral function with the zeros $0, \pm 1, \pm 2, \ldots$; we shall determine the extraneous factor e^{Gx} in this case.

Let us write

$$\sin \pi x = e^{Gx} fx,$$

where

$$fx = x \Pi' (1 - x/n) e^{x/n} \text{*}.$$

We have now to justify the process of logarithmic differentiation, not only for the present case but also for future factor-formulæ. Resuming the notation of § 111, we had the product

$$\prod_{n=r+1}^{\infty} \exp \{\operatorname{Log}(1 - x/a_n) + \phi_n (x/a_n)\},$$

which converged when $X \leqq R < A_{r+1}, A_{r+2}, \ldots$.

The polynomials ϕ_n were then chosen so that the series

$$\sum_{n=r+1}^{\infty} \{\operatorname{Log}(1 - x/a_n) + \phi_n (x/a_n)\}$$

converged; we could then replace the product of the exponentials by the exponential of the sum (§ 95). Thus inserting

* It is to be understood that n and m in this and the next two articles take negative as well as positive values, unless the contrary is explicitly stated.

again the factors corresponding to $a_1, a_2, \ldots, a_r$ we have for one of the values of the logarithm on the left-hand side,

$$\log \prod_{n=1}^{\infty} (1 - x/a_n) e^{\phi_n(x/a_n)} = \sum_{n=1}^{\infty} \{\operatorname{Log}(1 - x/a_n) + \phi_n(x/a_n)\}.$$

Thus the taking of logarithms is justified; to justify differentiation we must see that the conditions of the theorem of § 81 are fulfilled. The only point requiring proof is that the $\sum_{n=r+1}^{\infty}$ is uniformly convergent in (R). Taking the case when the function is of grade $k-1$, we have to prove that if $u_n = \sum_{p=k}^{\infty} x^p/pa_n^p$, then $\sum_{n=r+1}^{\infty} u_n$ is uniformly convergent in (R). Now

$$\begin{aligned} |u_n| &\leq \sum_{p=k}^{\infty} (X^p/pA_n^p), \\ &< \sum_{p=k}^{\infty} (X/A_n)^p, \\ &< \sum_{p=k}^{\infty} (R/A_n)^p, \\ &< \frac{R^k}{A_n^k(1 - R/A_n)}, = \alpha_n \text{ say}. \end{aligned}$$

This series $\Sigma\alpha_n$ has only positive terms and is convergent since by hypothesis $\Sigma 1/A_n^k$ is convergent. Thus the usual test for uniform convergence (§§ 73, 77) is satisfied. We can then by § 87 differentiate as often as we please.

Usually when the function has no grade, m_n is taken to be $n-1$; for this case the comparison with a series of positive terms involving R but not X was made in the last article.

We have then, in the present instance,

$$\log fx = \log x + \Sigma' \{\operatorname{Log}(1 - x/n) + x/n\},$$

and

$$f'x/fx = \frac{1}{x} + \Sigma' \left\{\frac{1}{x-n} + \frac{1}{n}\right\},$$

where the accent means that $n = 0$ is omitted.

Hence, by another differentiation, if $\psi x = f'x/fx$, we have

$$\psi' x = -1/x^2 - \Sigma' 1/(x-n)^2 = -\Sigma 1/(x-n)^2.$$

From the equation for $\sin \pi x$ we get

$$\pi \cot \pi x = G'x + \psi x,$$
$$-\pi^2/\sin^2 \pi x = G''x + \psi' x.$$

The function $G''x$ is either (1) an integral function (transcendental or rational), or (2) a constant. It will now be shown that the behaviour of $G''x$ when x tends to infinity makes (1) impossible.

I. Let $x, = \xi + i\eta$, tend to ∞ *by increasing values of* η, ξ remaining all the time between 0 and 1, the latter exclusive.

Since $$-\psi' x = \frac{1}{x^2} + \sum_{n=1}^{\infty}\left[\frac{1}{(\xi+i\eta+n)^2} + \frac{1}{(\xi+i\eta-n)^2}\right],$$
we have
$$|\psi' x| \leqq \frac{1}{|x|^2} + \sum_{n=1}^{\infty}\left[\frac{1}{|\xi+i\eta+n|^2} + \frac{1}{|\xi+i\eta-n|^2}\right].$$

Now $$1/|\xi+n+i\eta|^2 = 1/\{(\xi+n)^2+\eta^2\} < 1/(\eta^2+n^2),$$
$$1/|\xi-n+i\eta|^2 = 1/\{(\xi-n)^2+\eta^2\} < 1/(\eta^2+(n-1)^2),$$
and $$1/|x|^2 = 1/(\xi^2+\eta^2) < 1/\eta^2;$$
hence $$|\psi' x| < 2\sum_{n=0}^{\infty} 1/(\eta^2+n^2).$$

The series $\sum_{0}^{\infty} 1/(\eta^2+n^2)$ can be divided into two parts $\sum_{0}^{m-1}$ and $\sum_{m}^{\infty}$. The first of these can be made as small as we please, whatever be the value of m, by taking η sufficiently large; the second is less term by term than $\sum_{n=m}^{\infty} 1/n^2$, a quantity which can be made as small as we please by increasing m sufficiently. Hence $|\psi' x|$ tends to the limit zero when η tends to infinity.

It is easy to prove that $\pi^2/\sin^2 \pi x$ behaves in the same way. For

$$\sin \pi x = (e^{-\pi\eta}e^{\pi\xi i} - e^{\pi\eta}e^{-\pi\xi i})/2i,$$

and therefore tends to the limit ∞ when η tends to infinity.

Hence $$|G''x|, = |-\psi' x - \pi^2/\sin^2 \pi x|,$$
tends to zero when η tends to infinity. There exists therefore a real positive value η' such that when $|\eta| > \eta'$, we have, for $0 \leqq \xi < 1$,

$$|G''x| < \epsilon.$$

Inspection of the formulæ for $\psi' x$, $\pi^2/\sin^2 \pi x$ shows that each function has the period 1; hence $G''x$ has the period 1, and hence $|G''x| < \epsilon$ for *all* values of ξ when $|\eta| > \eta'$.

II. Next, let ξ tend to ∞, $|\eta|$ remaining less than or equal to η'. Since $G''x$ has the period 1, all values taken by $G''x$ within the strip contained between the straight lines $\eta = \pm \eta'$ are taken also in the rectangle bounded by $\eta = \pm \eta'$, $\xi = 0$, $\xi = 1$. As $G''x$ is finite for all finite values of x, its absolute value in this region must remain less than some definite number A; therefore also $|G''x| < A$ for the whole of the strip.

III. Since $\qquad |G''x| < A, \quad |G''x| < \epsilon,$

for points in and outside of the strip respectively, it follows that $G''x$ cannot be a rational or transcendental integral function (§§ 56, 93). *Hence $|G''x|$ is a constant; and the second inequality shows that this constant can only be zero.* Thus

$$G''x = 0,$$

and $$G'x = \text{a constant } b;$$

or $$\pi \cot \pi x = b + \frac{1}{x} + \Sigma' \left(\frac{1}{x-n} + \frac{1}{n} \right)$$

$$= b + \frac{1}{x} + \sum_1^\infty \frac{2x}{x^2 - n^2}.$$

Changing the sign of x and adding we have $b = 0$.

Hence

$$\pi \cot \pi x = \frac{1}{x} + \sum_{-\infty}^{\infty}{}' \left(\frac{1}{x-n} + \frac{1}{n} \right) \quad \text{..................(1)},$$

the formula for the cotangent in partial fractions.

Also $$Gx = \text{a constant } a;$$

hence $$\sin \pi x = e^a x \Pi' (1 - x/n) e^{x/n}.$$

Since $\lim\limits_{x=0} \dfrac{\sin \pi x}{x}$ is π, we have $e^a = \pi$. Hence finally we have, for all values of x, the factor-formula

$$\sin \pi x = \pi x \prod_{n=-\infty}^{\infty}{}' (1 - x/n) e^{x/n} \quad \text{..................(2)},$$

$$= \pi x \prod_{n=1}^{\infty} (1 - x^2/n^2) \quad \text{....................(2')}.$$

113. Formulæ for the other Circular Functions. A formula for $\cos \pi x$, analogous to (2), may be proved in the same way, or may be deduced from (2). Adopting the latter plan we can, most simply, write

$$\cos(\pi x/2) = \frac{\sin \pi x}{2\sin(\pi x/2)} = \frac{\Pi(1 - x^2/n^2)}{\Pi(1 - x^2/4n^2)} = \prod_{m=1}^{\infty}(1 - x^2/m^2),$$

where $m = 1, 3, 5, \dots$.

If we separate the factors of this expression into the simpler factors $1 \pm x/m$ we must, to ensure absolute convergence, attach to each an exponential factor. The product

$$\Pi(1 - x/m)\, e^{x/m},$$

where $m = \pm 1, \pm 3, \pm 5, \dots$, will be absolutely convergent, by the general theorem, and will, by pairing opposite values of m, give the preceding formula. Hence

$$\cos(\pi x/2) = \Pi(1 - x/m)\, e^{x/m} \quad\dots\dots\dots\dots(3),$$

where m takes all odd integer values.

By logarithmic differentiation we have

$$\tfrac{1}{2}\pi \tan(\pi x/2) = \Sigma\,[1/(m-x) - 1/m] \quad\dots\dots\dots\dots(4),$$

which may be deduced directly from (1) by the formula

$$2\cot \pi x = \cot(\pi x/2) - \tan(\pi x/2).$$

For the cosecant we have

$$2\csc \pi x = \cot(\pi x/2) + \tan(\pi x/2),$$

whence $$\pi \csc \pi x = \frac{1}{x} + \Sigma'\left(\frac{1}{x-2n} + \frac{1}{2n}\right) + \Sigma\left(\frac{1}{m-x} - \frac{1}{m}\right)$$

$$= 1/x + \Sigma'(-)^n\,[1/(x-n) + 1/n] \quad\dots\dots\dots\dots(5).$$

As a further illustration of factor-formulæ, let us resolve $\sin \pi x - \sin \pi a$ into factors, where a is any non-integral number. The zeros are $x = m - a$, and $x = 2n + a$, where m is an odd integer. As $\Sigma\, 1/(n-a)^2$ is absolutely convergent so are $\Sigma\, 1/(m-a)^2$ and $\Sigma\, 1/(2n+a)^2$. Hence

$$\sin \pi x - \sin \pi a = e^{Gx}\,\Pi\left(1 - \frac{x}{m-a}\right) e^{\frac{x}{m-a}}\, \Pi\left(1 - \frac{x}{2n+a}\right) e^{\frac{x}{2n+a}},$$

and the extraneous factor is determined precisely as for $\sin \pi x$

by showing that $G''x$ is a constant and that this constant is zero. We have then $G'x = b$,

$$\frac{\pi \cos \pi x}{\sin \pi x - \sin \pi a} = b + \Sigma \left(\frac{1}{x - m + a} + \frac{1}{m - a} \right)$$
$$+ \Sigma \left(\frac{1}{x - 2n - a} + \frac{1}{2n + a} \right);$$

and, when $x = 0$, $b = -\pi/\sin \pi a$.

Hence $\exp Gx = \exp(c - \pi x/\sin \pi a)$; and since, when $x = 0$, $-\sin \pi a = \exp G0 = \exp c$, we have finally

$$\exp Gx = -\sin \pi a \,.\, \exp(-\pi x/\sin \pi a).$$

In particular if $a = -1/2$ the system $m - a$ coincides with the system $2n + a$, each being of the form $(4n - 1)/2$; and we have

$$1 + \sin \pi x = e^{\pi x}\, \Pi \left[(1 - 2x/(4n - 1))\, e^{2x/(4n-1)}\right]^2,$$

or writing $x/2$ for x

$$1 + \sin(\pi x/2) = e^{\pi x/2}\, \Pi \left[(1 - x/(4n - 1))\, e^{x/(4n-1)}\right]^2.$$

Hence from (3)

$$\frac{1 + \sin(\pi x/2)}{\cos(\pi x/2)} = e^{\pi x/2} \frac{\Pi\, (1 - x/(4n - 1))\, e^{x/(4n-1)}}{\Pi\, (1 - x/(4n + 1))\, e^{x/(4n+1)}}.$$

Hence, by taking the derivate of the logarithm,

$$\tfrac{1}{2}\pi \sec(\pi x/2) = \pi/2 + \Sigma\, i^{m+1} \left(\frac{1}{x - m} + \frac{1}{m} \right) \quad \ldots\ldots(6),$$

where m is any odd integer, the factor i^{m+1} being $+1$ or -1 as m is of the form $4n - 1$ or $4n + 1$.

The formulæ (1, 4, 5, 6) effect the resolution of the elementary fractional circular functions into partial fractions. Instead of basing these on the factor-formula we might prove them independently.

For example, in the case of $\cot \pi x$, there is a simple infinity when $x = n$, but $\pi \cot \pi x - 1/(x - n)$ has no infinity when $x = n$. We cannot use the expression $\pi \cot \pi x - \Sigma\, 1/(x - n)$ since the sum is not convergent; we can however pair off opposite values of n and say that $\pi \cot \pi x - \sum_1^\infty [1/(x - n) + 1/(x + n)] - 1/x$ has no infinities in the finite part of the plane and is therefore Gx.

Or we can render $\Sigma\, 1/(x-n)$ absolutely convergent by adding to each term $1/n$ where n is not zero. Thus

$$\pi \cot \pi x - \Sigma' [1/(x-n) + 1/n] - 1/x$$

is at most a transcendental integral function; which is then shown to be zero.

This point of view leads to a very important theorem known as Mittag-Leffler's theorem, by which we can substitute for the simple infinities $a_1, a_2, \ldots, a_n, \ldots$, where $\mathrm{L}\, a_n$ is ∞, isolated singular points of the most general kind which a one-valued function can have, and then write down an expression for such a function. For this theorem we must refer to more advanced treatises, as we shall not require here more than can be deduced at once from the factor-theorem.

Ex. 1. Let $fx = \Pi(1 - x^2/m^2)$ where m is odd. Prove without appeal to known properties of the cosine that $f(x+4) = fx$.

Ex. 2. Deduce (3) from (2) by using the formula

$$\sin \pi (x + 1/2) = \cos \pi x.$$

Ex. 3. Deduce (4) from (1) by the formula

$$\cot \pi (x + 1/2) = -\tan \pi x.$$

Ex. 4. In (1) write for x, $=\xi + i\eta$, the conjugate number $\bar{x}$, $=\xi - i\eta$, and then subtract the formula so obtained from (1). We obtain

$$\pi (\cot \pi\bar{x} - \cot \pi x) = (x - \bar{x}) S,$$

where $S = \Sigma \dfrac{1}{(x-n)(\bar{x}-n)}$ = the sum of the squares of the reciprocals of the distances between (ξ, η) and the points $0, \pm 1, \pm 2, \ldots$. Hence prove that

$$S = \frac{\pi}{\eta} \frac{\sinh 2\pi\eta}{\cosh 2\pi\eta - \cos 2\pi\xi}.$$

Ex. 5. Comparing (2′) with the known power series for $\sin \pi x$, prove that

$$\sum_1^\infty 1/n^2 = \pi^2/6, \quad \sum_1^\infty \sum_1^\infty 1/n_1^2 n_2^2 = \pi^4/5!.$$

114. The Gamma Function and its Reciprocal. We can also write down the general form of a transcendental integral function with simple zeros at the points $-1, -2, -3, \ldots$, namely

$$e^{Gx} \prod_{n=1}^{\infty} (1 + x/n) e^{-x/n}.$$

If we suppose the extraneous factor to be $e^{\gamma x}$, where γ is Euler's constant $\mathrm{L}(1 + 1/2 + \ldots + 1/n - \mathrm{Log}\, n)$, (§ 109), we obtain

a function of considerable importance, which Weierstrass names the *Factorielle* of x, and writes $\mathrm{Fc}\,x$; Gauss denoted this same function by $1/\Pi x$.

We have, as the definition of $\mathrm{Fc}\,x$,

$$\begin{aligned}\mathrm{Fc}\,x &= \mathrm{L}\exp(x+x/2+\ldots+x/n-x\,\mathrm{Log}\,n)(1+x)e^{-x}\\ &\qquad\qquad .(1+x/2)e^{-x/2}\ldots(1+x/n)e^{-x/n}\\ &= \mathrm{L}\,\frac{(1+x)(2+x)\ldots(n+x)}{1.2\ldots n}n^{-x}.\end{aligned}$$

Gauss expressed Πx by the inverse of this latter form.

If we write $x-1$ for x we have

$$\mathrm{Fc}\,(x-1)=\mathrm{L}\,\frac{x(1+x)\ldots(n+x-1)}{1.2\ldots n-1}n^{-x}.$$

Hence $$\mathrm{Fc}\,x/\mathrm{Fc}\,(x-1)=\mathrm{L}\,\frac{n+x}{nx}=1/x,$$

or $x\,\mathrm{Fc}\,x=\mathrm{Fc}\,(x-1)$; while $\mathrm{Fc}\,0=1$.

The reciprocal function, namely $1/\mathrm{Fc}\,x$, is Euler's gamma function $\Gamma(1+x)$, defined otherwise in the Infinitesimal Calculus, for a real x greater than -1, by the integral

$$\Gamma(1+x)=\int_0^\infty e^{-t}t^x\,dt.$$

F. W. Newman* published in 1848 (*Camb. and Dub. Math. J.* vol. iii. p. 59) a memoir on the gamma function in which use is made of the formula

$$\Gamma x=\frac{e^{-\gamma x}}{x}\cdot\frac{e^x}{1+x}\cdot\frac{e^{x/2}}{1+x/2}\cdots\cdots;$$

this formula implies the use of primary factors for $1/\Pi x$.

$\Gamma(1+x)$ is a transcendental fractional function of x with simple infinities at -1, -2, -3,...; by a mere change of notation we see that Γx has the following properties:

$$\Gamma(1+x)=x\Gamma x,\ \Gamma 1=1.$$

From the formula

$$\frac{\sin\pi x}{\pi x}=\Pi'(1-x^2/n^2),$$

* Gauss was the first to express $1/\Pi x$ as an infinite product of factors of the type $(1+x/n)\exp\lambda_n x$ (λ_n independent of x), each of which vanishes at one and only one zero of the function. Historically it was the consideration of this special function that led Weierstrass to the general factor-theorem of this chapter. Gauss's symbol Π is much used in English works; but it is liable to confusion with the universal notation for products.

we have at once

$$\frac{\sin \pi x}{\pi x} = \mathrm{Fc}\, x\, \mathrm{Fc} - x,$$

or

$$\pi x/\sin \pi x = \Gamma(1+x)\,\Gamma(1-x);$$

whence in particular for $x = 1/2$,

$$\pi/2 = \Gamma(3/2)\,\Gamma(1/2),$$

or

$$\pi = \{\Gamma(1/2)\}^2;$$

taking the square roots and observing that Γx is, from its definition, positive for a positive x, we have

$$\Gamma(1/2) = \pi^{1/2}.$$

Ex. 1. $\operatorname{Lim}_{q=\infty} \prod_{n=q}^{n=pq} (1+x/n) = p^x.$

Ex. 2. Observing that $\mathrm{L}\,(1-t/n)^n$ is e^{-t}, prove that

$$\int_0^\infty e^{-t} t^x dt = \mathrm{L} \frac{1 \cdot 2 \ldots n}{(1+x)(2+x)\ldots(n+x)}\, n^x,$$

where x is complex.

Ex. 3. Prove that $\operatorname{Log} \Gamma(x+1) = -\gamma x + \frac{1}{2}s_2 x^2 - \frac{1}{3}s_3 x^3 + \frac{1}{4}s_4 x^4 - \ldots$, where $|x| < 1$ and $s_r = \sum_{n=1}^{\infty} 1/n^r$

From Weierstrass's expression for $1/\Gamma(x+1)$ as a product of primary factors, we can deduce at once the formula

$$D^2 \log \Gamma x = \frac{1}{x^2} + \frac{1}{(x+1)^2} + \frac{1}{(x+2)^2} + \ldots\ldots + \frac{1}{(x+n)^2} + \ldots\ldots$$

Calling this function ψx, we have

$$\psi(x+1) - \psi x = -1/x^2;$$

$$\psi(1-x) + \psi x = \Sigma\, 1/(x+n)^2 = \pi^2 \csc^2 \pi x.$$

From the properties of ψx it is possible to evolve those of Γx; this has been done by Hermite.

Ex. 1. Prove that

$$\psi x + \psi(x+1/n) + \ldots + \psi(x+(n-1)/n) = n^2 \psi(nx).$$

Ex. 2. Prove that

$$\frac{\Gamma a\, \Gamma x}{\Gamma(a+x)} = \frac{1}{x} + \sum_{n=1}^{x} \frac{(-)^n (a-1)(a-2)\ldots(a-n)}{1 \cdot 2 \ldots n} \frac{1}{x+n}.$$

CHAPTER XVI.

INTEGRATION.

115. Definitions of an Integral. There are two available definitions for a definite integral $\int_\alpha^\beta f\xi d\xi$, where $f\xi$ is a function of a real variable ξ. One of these treats the definite integral as the limit of a sum of elements $f\xi\, d\xi$, the other treats integration as a process inverse to differentiation. These definitions, which we shall call the *first* and *second definitions*, can be generalized so as to apply to analytic functions of x; we shall discuss both definitions and show that they lead to the same results. Let us begin with the simple example $\int_{x_0}^x x^n dx$, where n is a positive integer, and use the first definition.

First we are to think of a path from x_0 to x; we shall call x_0, x *end-values* of the path, in preference to the misleading term *limits*. On this path we are to interpolate μ points $x_1, x_2, \ldots, x_\mu$ so that we have $\mu+2$ points $x_0, x_1, x_2, \ldots, x_{\mu+1}$, where $x_{\mu+1}$ is merely the final value x itself. The points are to be so interpolated that by increasing μ sufficiently, all the strokes $x_{\lambda+1}-x_\lambda$, where $\lambda=0, 1, 2, \ldots, \mu$, can be made as small as we please; the points may for example be at equal distances, or may divide the path, which we suppose to possess length, into equal arcs. By $x^n dx$ we understand a typical expression for $x_\lambda{}^n(x_{\lambda+1}-x_\lambda)$, and by the integral itself the limit (if there is one) of the sum $\sum_{\lambda=0}^{\mu} x_\lambda{}^n(x_{\lambda+1}-x_\lambda)$, when μ tends to infinity.

The novelty lies in the infinity of possible *paths of integration* between x_0 and x; in the case of the real variable there are but two, namely along the real axis between x_0 and x, or along the rest of the real axis viâ ∞.

Now we have

$$x_{\lambda+1}^{n+1} - x_\lambda{}^{n+1} = (x_\lambda + h_\lambda)^{n+1} - x_\lambda{}^{n+1},$$

where $h_\lambda = x_{\lambda+1} - x_\lambda$. Hence

$$x_{\lambda+1}^{n+1} - x_\lambda{}^{n+1} = (n+1)\, h_\lambda\, (x_\lambda{}^n + \epsilon_\lambda) \;\ldots\ldots\ldots\ldots (1),$$

where ϵ_λ can be made as small as we please by making h_λ small enough.

Let ϵ be the greatest of the numbers ϵ_λ where $\lambda = 0, 1, 2, \ldots, \mu$; then ϵ is also as small as we please, if we take μ large enough and therefore all the numbers h_λ small enough.

Adding the $\mu + 1$ equations (1) we have

$$x_{\mu+1}^{n+1} - x_0{}^{n+1} = (n+1)\,\Sigma\, x_\lambda{}^n h_\lambda + (n+1)\,\Sigma\,\epsilon_\lambda h_\lambda;$$

or

$$\begin{aligned} \left| \frac{x_{\mu+1}^{n+1} - x_0{}^{n+1}}{n+1} \quad \Sigma x_\lambda{}^n h_\lambda \right| &= \left| \Sigma\, \epsilon_\lambda h_\lambda \right| \\ &\leqq \Sigma \left| \epsilon_\lambda \right| \left| h_\lambda \right| \\ &\leqq \epsilon\, \Sigma \left| h_\lambda \right|. \end{aligned}$$

But when μ tends to ∞, ϵ tends to 0 and $\Sigma\,|h_\lambda|$ to the length of the path. Thus, for a path of finite length,

$$\lim_{\mu=\infty} \Sigma\, x_\lambda{}^n\, h_\lambda = \frac{x_{\mu+1}^{n+1} - x_0{}^{n+1}}{n+1},$$

or

$$\int_{x_0}^{x} x^n dx = \frac{x^{n+1} - x_0{}^{n+1}}{n+1} \;\ldots\ldots\ldots\ldots\ldots\ldots (2).$$

In this case the integral is independent of the path, but this is merely owing to the simplicity of the instance; we have seen already that $\int_{x_0}^{x} dx/x$ does depend on the path from x_0 to x, and this is a fact of cardinal importance. It will be observed that in this case when n is -1 the right-hand side of (2) is meaningless; for all other integer values of n the equation (2) holds good, and the proof is the same, except that we appeal to the binomial

theorem for a negative exponent; but when n is a negative integer we must exclude paths which pass through the origin, for the origin is an infinity of x^n.

Now let us consider the equation (2) in connexion with the second definition of integration. The function x^n will be replaced, for the sake of greater generality, by the power series $P(x-c)$ and x_0, x will be supposed to lie in the domain of this power series. By the indefinite integral of

$$P(x-c) = a_0 + a_1(x-c) + a_2(x-c)^2 + \dots,$$

is to be understood the new power series $\bar{P}(x-c)$ defined by

$$\bar{P}(x-c) = \int P(x-c)\,dx = \text{const.} + a_0(x-c) + a_1\frac{(x-c)^2}{2} + a_2\frac{(x-c)^3}{3} + \dots. \quad (\S\ 86).$$

Since $D\bar{P}(x-c) = P(x-c)$, the series $P(x-c)$ and $\bar{P}(x-c)$ have the same domain (§ 87). We now define $\int_{x_0}^{x} P(x-c)\,dx$ as the value of $\bar{P}(x-c)$ at x minus the value of the same series at x_0, that is as

$$\bar{P}(x-c) - \bar{P}(x_0-c).$$

We shall consider presently what modification must be introduced into this second definition when x_0 and x do not both belong to the domain mentioned above. For the moment we are concerned only with what happens when $P(x-c)$ is converted into a terminating series x^n, with an infinite domain, by making $c=0$, $a_n = 1$, and all the coefficients a_r, except a_n, vanish. In this case $\bar{P}(x-c)$ becomes const. $+\dfrac{x^{n+1}}{n+1}$, and $\bar{P}(x-c) - \bar{P}(x_0-c)$ becomes $\dfrac{x^{n+1} - x_0^{\,n+1}}{n+1}$. Thus equation (2) holds whichever definition we adopt for integration; in the one case the passage is made from x_0 to x by a path of integration, in the other use is made of the analytic function const. $+x^{n+1}/(n+1)$ and the passage from x_0 to x is made in one stride.

116. Reconciliation of the Definitions in the case of the Power Series. The instance of the preceding paragraph

shows what is meant by the integral of a function fx taken along a given path of finite length, when the integral is treated as the limit of a sum. We suppose the function to be one-valued and continuous along the path,—this being in all cases which concern us a limitation on the path selected, not on the function; we interpolate as before μ points on the path between the initial point x_0 and the final point x, and we seek

$$\lim_{\mu=\infty} \sum_{\lambda=0}^{\mu} fx_\lambda (x_{\lambda+1} - x_\lambda).$$

This limit, if it is finite, is the integral of the function along that path according to the first definition and is denoted by $\int_{x_0}^{x} fx\,dx$, or $\int_L fx\,dx$, where L is the path of integration.

Instead of discussing the integral of a function no matter how given, we shall suppose as elsewhere in this book that fx is analytic. We have to reconcile the two definitions by showing that *for the first definition* $\int_{x_0}^{x} fx\,dx$, for $fx = P(x-c)$ and for a path x_0 to x lying wholly within the domain of $P(x-c)$, is equal to $\bar{P}(x-c) - \bar{P}(x_0 - c)$. For simplicity we put $c = 0$ and write $Px = \Sigma a_n x^n$.

Clearly we shall be able to accomplish our purpose if we can show that

$$\int_{x_0}^{x} (\Sigma a_n x^n)\,dx = \Sigma \int_{x_0}^{x} a_n x^n\,dx\,;$$

for then the right-hand side can be replaced by

$$\Sigma a_n (x^{n+1} - x_0^{n+1})/(n+1)\,;$$

that is, by

$$\bar{P}x - \bar{P}x_0.$$

Account has to be taken of the combination of two infinite processes leading to a double limit.

We suppose R the radius of convergence; and we suppose the path of integration L to lie within (R). Then we know that Px is uniformly convergent; that is, denoting by $r_n x$ the remainder after n terms of the series, we can choose n so that, throughout

L, $|r_n x| < \epsilon$, where ϵ is an arbitrary positive number. We have, when n is so chosen,

$$\int_L Px\,dx = \int_L \left(\sum_0^{n-1} a_p x^p + r_n x\right) dx$$
$$= \sum_0^{n-1} \int_L a_p x^p\,dx + \int_L r_n x\,dx.$$

But
$$\left|\int_L r_n x\,dx\right| \leq \int_L |r_n x\,dx|$$
$$< \int_L \epsilon\,|dx|$$
$$< \epsilon \times \text{length of } L.$$

Hence the difference between the integral of the sum of the series, and the sum of the integrals of the first n terms, can be made as small as we please by taking n large enough. That is,

$$\int_L \sum_0^\infty a_n x^n\,dx = \sum_0^\infty \int_L a_n x^n\,dx \ldots\ldots\ldots\ldots\ldots\ldots(1).$$

Let the path L start from the centre o, then the above theorem says that

$$\int_0^x Px\,dx = \sum_0^\infty a_n x^{n+1}/(n+1) \ldots\ldots\ldots\ldots\ldots (2),$$

a new power series without a constant term which we denote by $\bar{P}x$.

It must be carefully noticed that the path of integration lies altogether in the circle of convergence, but with this restriction it matters not what the path may be, for $\bar{P}x$ is a one-valued and continuous function of x as long as x keeps within the circle. Thus not only is our integral $\int_L Px\,dx$ proved to exist for a path L in (R), but further its value is independent of the path,—the end-values being given.

If the path lead from x_0 to x, then from (1)

$$\int_{x_0}^x Px\,dx = \sum_0^\infty a_n \frac{x^{n+1} - x_0^{n+1}}{n+1}$$
$$= \bar{P}x - \bar{P}x_0.$$

Hence the integral from x to x_0 is the opposite of the integral from x_0 to x; that is, if we reverse the path we change the sign of

the integral. And further, by making $x=x_0$, we see that *the integral along any closed path lying in* (R) *is* o.

The above argument serves to show more generally that if $\sum_{n=1}^{\infty} f_n x$ is a series of functions which are one-valued and continuous in a region Γ, and if the series is uniformly convergent in Γ, then for a path of integration L which lies in Γ and is of finite length,

$$\int_L \Sigma f_n x\, dx = \Sigma \int_L f_n x\, dx.$$

It will be observed that the theorem is still true if the convergence be given as uniform merely along L. For our purposes it is sufficient to take functions which are analytic and series which are uniformly convergent over a region Γ.

Closely related to this theorem on the integration term by term of a series of analytic functions is one on the differentiation term by term of such series. Suppose that $\Sigma f'_n x$ is uniformly convergent in Γ (the terms $f_n x$ being defined as before), then $\Sigma f_n x$ (which can be proved to be analytic over Γ) has $\Sigma f'_n x$ for its derivate. For $\int_L \Sigma f'_n x\, dx = \Sigma \int_L f'_n x\, dx = \Sigma (f_n x - f_n x_0)$, where x_0, x are the initial and terminal points of L; and differentiation with respect to x gives

$$\Sigma f'_n x\, dx = \frac{d}{dx} \Sigma f_n x.$$

We may add that by successive integrations along a suitable path L of a series of power series of the kind considered in § 81, we get a theorem which is for integration the exact analogue of that of § 87 for differentiation.

117. Case where the End-values belong to Different Elements. (1) Let us take the case where x_0 lies in the domain of $P(x-c_0)$, x in the domain of $P(x-c_0|c_1), = Q(x-c_1)$, an immediate continuation of $P(x-c_0)$. Replacing in these series $(x-c)^n$ by $(x-c)^{n+1}/(n+1)$, we get new series which have the original series for first derivates. Denote by $\bar{P}(x-c_0)$ the series $\sum_{n=0}^{\infty} a_n (x-c_0)^{n+1}/(n+1)$ which has no constant term; the continuation of this series into the second domain is obtained directly from Taylor's theorem and is

$$\begin{aligned}
\bar{P}&(c_1-c_0+x-c_1) \\
&= \bar{P}(c_1-c_0) + P(c_1-c_0)(x-c_1) + P'(c_1-c_0)(x-c_1)^2/2! + \ldots \\
&= \bar{P}(c_1-c_0) + \text{the series which arose from } Q(x-c_1) \text{ by} \\
&\qquad\qquad \text{replacing } (x-c_1)^n \text{ by } (x-c_1)^{n+1}/(n+1).
\end{aligned}$$

Let us call this continuation $\bar{Q}(x-c_1)$.

The meaning to be assigned to the integral $\int_{x_0}^{x} fx\,dx$ if the second definition be adopted is therefore this:—

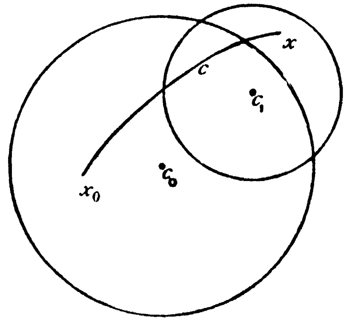


Fig. 50.

Let $\int_{x_0}^{x} fx\,dx$ mean

$$\int_{x_0}^{c} P(x-c_0)\,dx + \int_{c}^{x} Q(x-c_1)\,dx,$$

where c is in the common domain of P and Q; and let us evaluate the integrals in these two cases by using

$$\bar{P}(x-c_0),\ \bar{Q}(x-c_1)$$

respectively. The sum is

$$\bar{Q}(x-c_1) - \bar{Q}(c-c_1) + \bar{P}(c-c_0) - \bar{P}(x_0-c_0),$$

or

$$Q(x-c_1) - P(x_0-c_0);$$

since $\bar{Q}(x-c_1)$, $\bar{P}(x-c_0)$ agree in their common domain. It should be observed that c has dropped out from the final formula.

If we denote by Fx that portion of an analytic function which is defined by $\bar{P}(x-c_0)$ and $\bar{Q}(x-c_1)$, the passage from x_0 to x is effected in two steps, namely x_0 to c, c to x, and

$$\int_{x_0}^{x} fx\,dx = Fx - Fc + Fc - Fx_0 = Fx - Fx_0.$$

The same result flows from the first definition, the only difference is that the steps from x_0 to c, c to x of Fx are replaced by the integrals of $P(x-c_0)$, $Q(x-c_1)$ over paths from x_0 to c, c to x, which lie wholly within the domains of $P(x-c_0)$, $Q(x-c_1)$ as in the figure.

(2) When more than two elements are needed to get from x_0 to x the process is as follows: continue $P(x-c_0)$ by a standard chain of series $P_1(x-x_1), \ldots, P_n(x-x_n)$, $Q(x-c_1)$ until a series $Q(x-c_1)$ is reached which contains x in its domain. The passage from x_0 to x can be effected by a path $x_0c'c'' \ldots c^{(n)}c^{(n+1)}x$ such that the parts x_0c', $c'c''$, $\ldots$, $c^{(n)}c^{(n+1)}$, $c^{(n+1)}x$ lie wholly in the

domains of the elements about $c_0, x_1, x_2, \ldots, x_n; c_1$. Then $\int_{x_0}^{x} fx\,dx$ *for this path and these elements* is defined

(i) as $(Fc' - Fx_0) + (Fc'' - Fc') + \ldots + (Fc^{(n+1)} - Fc^{(n)})$

$$+ (Fx - Fc^{(n+1)}) = Fx - Fx_0,$$

where Fx is that part of an analytic function which arises from $\bar{P}(x - c_0)$ by continuation along the given chain;

(ii) as the sum of the integrals (treated as limits of sums) of $P(x - c_0)$, $P_1(x - x_1)$, $P_2(x - x_2), \ldots$, $P_n(x - x_n)$, $Q(x - c_1)$ along paths x_0 to c', c' to c'', c'' to c''', $\ldots$, $c^{(n)}$ to $c^{(n+1)}$, $c^{(n+1)}$ to x that lie wholly within the domains of the corresponding series.

As in the case of two elements it is indifferent whether we use the first or second definition.

In the following theorems,—which are direct consequences of either definition,—L is a path about every point of which fx is analytic:—

I. *The integral along L from x_0 to b is equal to the sum of the integrals along L from x_0 to c and from c to b, where c is any point of L.*

II. *The integrals $\int_{x_0}^{x} fx\,dx$ and $\int_{x}^{x_0} fx\,dx$ taken along L from x_0 to x and from x to x_0 respectively are equal in magnitude and opposite in sign.*

III. *The integral $\int_{x_0}^{x} A fx\,dx$ is equal to $A\int_{x_0}^{x} fx\,dx$, where A is a constant and the path from x_0 to x is the same for both integrals.*

IV. *The integral $\int_L (f_1x + f_2x + \ldots + f_nx)\,dx$, where the path of integration is L, is equal to the sum of the integrals $\int_L f_1x\,dx$, $\int_L f_2x\,dx, \ldots, \int_L f_nx\,dx$.*

The natural generalization of IV. is to replace the sum of the n f's by an infinite series which is uniformly convergent over L.

If we wish to change the variable x in the equation $Fx = \int fx\,dx$

by a substitution $x=\phi y$, where ϕy is a one-valued and analytic function, the result is $F\phi y=\int f\phi y\,.\,\phi' y dy$ as for a real variable.

For by differentiation we get

$$F'\phi y\,.\,\phi' y=f\phi y\,.\,\phi' y,$$

that is, $$F'\phi y=f\phi y,$$

which is right since $$F'x=fx.$$

An examination of the definite integral shows that the value of that integral may depend on the chain or path from x_0 to x, since Fx was defined by means of a particular chain of series, and therefore Fx in $Fx-Fx_0$ may be altered in value by the employment of a new chain or path originating from the element about c_0. An illustration is afforded by the formula

$$\int_0^x \frac{dx}{1+x^2}=\tan^{-1}x.$$

Here the integral of a one-valued function is a many-valued function. The formula only receives its full meaning when account is taken of the infinite multiplicity of paths from 0 to x.

The general question suggested is this: *given an analytic function fx and two paths L, L' from x_0 to x, is the value of $\int_{x_0}^{x} fx\,dx$ for the path L equal to or connected by a simple relation with the value of the same integral for the path L'?*

Suppose that fx under the sign of integration is replaced by an element of an analytic function, and that x_0hx, x_0kx are two paths lying wholly within the domain of this element; then denoting by (x_0hx), (x_0kx), (xkx_0) the integrals from x_0 to x along the paths x_0hx, x_0kx, and from x to x_0 along xkx_0, we have proved (§ 116) that $(x_0hx)=(x_0kx)$, an equation which can be transformed into

$$(x_0hx)+(xkx_0)=0\,;$$

this means that the integral over the closed path x_0hxkx_0 is zero. Cauchy discovered a general theorem relating to the vanishing of $\int fx\,dx$ when taken over a closed path; or,—replacing such an equation as $(x_0hx)+(xkx_0)=0$ by $(x_0hx)=(x_0kx)$,—to the possibility of deforming one path from x_0 to x into another from x_0 to x without altering the value of the associated integral.

The regions of which we shall speak in the subsequent articles of this chapter will be assumed not to contain the point ∞ unless the contrary is explicitly stated.

118. Cauchy's Theorem. *Given that fx is analytic over a simply connected region Γ, then over any circuit C in Γ,*

$$\int_C fx\,dx = 0.$$

Consider the analytic function

$$Fx = \int_{x_0}^{x} fx\,dx,$$

where x_0 is some fixed point of Γ and the path from x_0 to x lies wholly in Γ. Corresponding to each element of ϕx (§ 107) there is one of Fx of equal domain; hence since the radii of the elements of ϕx have a lower limit which is greater than zero, so also have the elements of Fx. This means that Fx has no singular points in Γ, and gives therefore for that region a localized one-valued function. Now $\int_C fx\,dx$ by the first definition is merely the change of Fx taken over the circuit C; hence applying the theorem of § 107 we have at once

the change of Fx over $C = 0$,

and the theorem is established.

In a sense Cauchy's theorem states nothing more than is contained in the earlier paragraph to which we have referred. For Fx is an analytic function equally with fx, the only difference being that, whereas fx is evolved from some element $\Sigma a_n(x-x_0)^n$, Fx arises from const. $+ \Sigma a_n(x-x_0)^{n+1}/(n+1)$. It is to be observed however that Fx is not considered by itself, but in close conjunction with fx; for the singular points of Fx are singular points of fx, though possibly of a new kind as in the case of $1/x$ and its indefinite integral const. $+ \log x$. When fx is one-valued the change of fx over any circuit is zero, even though the circuit contain singular points; but it does not follow that Fx will be one-valued and therefore the change of Fx over this circuit may differ from zero. The cause of this phenomenon will appear in

the next article. But first we must complete Cauchy's theorem by taking a region in the finite part of the plane which is bounded by non-intersecting circuits $C, C_1, C_2, \ldots, C_n$ such that $C_1, C_2, \ldots, C_n$ are all contained in C as in fig. 51. We have to show that when fx is one-valued and analytic over the region,

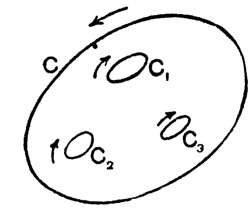


Fig. 51.

$$\int_C fx\,dx + \int_{C_1} fx\,dx + \ldots + \int_{C_n} fx\,dx = 0,$$

when the curves are described positively with regard to the region.

Draw lines pq, $p'q'$, ... as in fig. 52 (i) and treat the resulting figure as a limiting case of fig. 52 (ii). The shaded region is

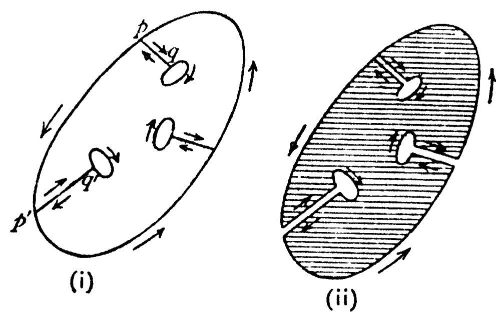


Fig. 52.

now bounded by a single closed curve; moreover when this closed curve is described positively each of the lines pq, $p'q'$, ... is described twice in opposite directions. As

$$\int_p^q fx\,dx + \int_q^p fx\,dx = 0,$$

with similar equations for $p'q'$, ..., all the integrals due to the new lines disappear and the remaining integrals combine to give

$$\int_C fx\,dx + \int_{C_1} fx\,dx + \ldots + \int_{C_n} fx\,dx = 0.$$

119. Residues. *When the function fx is as before, except that Γ contains a finite number of non-essential singular points $c_1, c_2, \ldots, c_\lambda$ of fx, the integral*

$$\int_C fx\,dx = 2\pi i \Sigma a_{-1},$$

where the integral is taken positively over the curves that constitute the boundary C of Γ, and a_{-1} is the coefficient of $1/(x-c)$ in the expansion $\sum_{n=-m}^{\infty} a_n(x-c)^n$ at a singular point c.

Since the integral of the sum of a series with a finite number of terms (say $n+1$) equals the sum of the integrals of those terms, it follows that if $\int_{(c)}$ denotes integration in a positive sense over the contour of a small circular region with centre c, then

$$\int_{(c)} \sum_{n=-m}^{\infty} a_n (x-c)^n dx = a_{-m}\int_{(c)}\frac{dx}{(x-c)^m}+\dots$$
$$+a_{-1}\int_{(c)}\frac{dx}{x-c}+\int_{(c)}P(x-c)\,dx,$$
$$=a_{-1}\int_{(c)}\frac{dx}{x-c}=2\pi i a_{-1} \quad (\S\ 30),$$

because all the other terms in the series of integrals vanish. Here we make use of $\int_{(c)} x^m dx = 0$ when the integer $m \neq -1$, and of $\int Pxdx = 0$ when the integration extends over a circuit in the domain of Px.

Let all the points $c_1, c_2, \dots, c_\lambda$ be made the centres of small circles and let Γ' be bounded by the curves C and by these small circles. Since fx satisfies all the requirements of Cauchy's theorem for this region Γ' which has arisen from Γ by cutting out the c's, it follows that

$$\int_C fxdx - \int_{(c_1)} fxdx - \int_{(c_2)} fxdx - \dots - \int_{(c_\lambda)} fxdx = 0,$$

where the circles of centres $c_1, c_2, \dots, c_\lambda$ are described positively with respect to their centres. Hence

$$\int_C fxdx = 2\pi i \Sigma a_{-1}.$$

This theorem shows that the coefficient a_{-1} and the term $a_{-1}(x-c)^{-1}$ in an expansion $\sum_{n=-m}^{\infty} a_n(x-c)^n$ play a specially important rôle. Cauchy has given to a_{-1} the name *residue relative to c, or at c.* At non-singular finite points there is of course no residue; but the point ∞ needs special examination.

When ∞ is a non-essential singular point, the expansion about ∞ is $\sum_{n=m}^{-\infty} a_n x^n$, where m is a positive integer; when ∞ is a non-singular point we have the expansion $\sum_{n=0}^{-\infty} a_n x^n$; thus we have a term a_{-1}/x even when ∞ is a non-singular point. By the integral round ∞ we understand the integral in the negative sense (that is, the positive sense with regard to ∞) round a circle so large as to include all singular points other than the possible singular point $x=\infty$ (supposed isolated); thus

$$\int_{(\infty)} fx dx = -2\pi i a_{-1}.$$

The residue at ∞ is defined to be $-a_{-1}$.

For one-valued analytic functions we have the following general theorems:—

The integral of a one-valued analytic function round a circuit which contains only non-essential singular points, and of these only a finite number, is $2\pi i$. (*the sum of the residues at these points*).

Similarly we can prove that *if the only singular points outside the circuit are non-essential, and are finite in number, the value of the same integral is* $-2\pi i$. (*the sum of the residues at these points*).

Hence it follows that *the sum of all the residues of a rational fraction is zero.*

The inference is that *an integral from* x_0 *to* x *will alter its value, in general, when the path is deformed continuously from* x_0hx *to* x_0kx, *if it pass over singular points in the process.*

When a function is one-valued and analytic, the value of $\int fx dx$ *taken positively over a closed curve* A *is unaffected by any continuous deformation of* A *which does not involve a passage over singular points.*

For let A, B be two positions of the moving circuit. By Cauchy's theorem

$$\int_A fx dx + \int_B fx dx = 0,$$

where the second integral is described *negatively* with respect to the interior of B.

In particular when a circuit can be made to vanish by a continuous contraction which does not require a passage over a singular point of the function, the value of the integral of the function taken over the circuit is zero.

120. General Applications of the Theory of Residues. Suppose that fx is one-valued and analytic about all points of a region Γ bounded by one or several closed curves, as in fig. 51, and that c is a point in Γ. The integral $\frac{1}{2\pi i}\int_A \frac{fx\,dx}{x-c}$, taken over a circuit A in Γ, may be called Cauchy's integral, for it plays an essential part in the development of the theory of functions along Cauchy's lines. We shall assume that c is not a zero of fx, and that A can be contracted continuously until it vanishes, without passing out of Γ.

Within the region bounded by A, the function

$$fx/(x-c)$$

has only one singularity, namely an infinity of the first order at c. The residue at this point is found at once by the use of Taylor's theorem; for the element at c of the analytic function fx is

$$fc + f'c\,(x-c) + f''c\,(x-c)^2/2\,! + \dots,$$

and therefore the residue in question is fc. But

$$\begin{aligned}\int_C fx\,dx/(x-c) &= 2\pi i\,\{\text{the residue of } fx/(x-c) \text{ within } A\}\\ &= 2\pi i\, fc;\end{aligned}$$

hence we deduce this very useful theorem, due to Cauchy:

If fx be an analytic function of x which is one-valued and analytic about the points on and interior to a circuit A of the x-plane and if c be a point interior to A, then

$$fc = \frac{1}{2\pi i}\int_A \frac{fx\,dx}{x-c},$$

the integral being taken positively over the circuit A.

Ex. Prove that if $x-c$ be replaced by $(x-c_1)(x-c_2)\dots(x-c_\lambda)$, then

$$\frac{1}{2\pi i}\int_A \frac{fx\,dx}{(x-c_1)(x-c_2)\dots(x-c_\lambda)} = \sum_{\mu=1}^{\lambda} \frac{fc_\mu}{f'c_\mu}\,\frac{1}{x-c_\mu}.$$

If about a point c, $fx = (x-c)^{-m} P_0(x-c)$, then c is an infinity of order m. If m be a negative integer, then c is a zero of order m. Let then m be an integer positive or negative; and let

$$fx = (x-c)^m P_0(x-c).$$

Then
$$\log fx = m\log(x-c) + \log P_0(x-c),$$
$$= m\log(x-c) + P_0(x-c),$$

where the circle of convergence of the new power series extends to a zero or singular point of fx. Hence

$$f'x/fx = m/(x-c) + P_0(x-c).$$

Let gx be a function which is analytic over a region Γ which contains a finite number of zeros and isolated singular points of fx, the latter being all non-essential; then

$$gx = gc + g'c(x-c) + \ldots,$$

where c is a point of Γ; hence when c is a zero of fx whose order is m, we have

$$gx\, f'x/fx = mgc/(x-c) + P_0(x-c);$$

and the residue of $gx\,f'x/fx$ at c is mgc. Similarly if c' be an infinity of order m' lying within Γ, the residue of $gx\,f'x/fx$ at c' is $-m'gc'$. Hence if A be a circuit lying in the region Γ,

$$\frac{1}{2\pi i}\int_A gx\, d\log fx = \Sigma(mgc - m'gc'),$$

the summation applying to all zeros c and infinities c' of fx which lie within A; it being of course understood that when c is a zero its order is m, but when c is an infinity its order is m'. When gx is a constant we have again a result already considered (§ 108). When $gx = x$ we have the following theorem:—

$$\frac{1}{2\pi i}\int_A x\frac{f'x}{fx}dx = \Sigma mc - \Sigma m'c',$$

that is, *the integral is equal to the sum of the values for which fx vanishes diminished by the sum of the values for which fx is infinite, multiple zeros and infinities being counted as often as is indicated by their orders of multiplicity.*

121. Special Applications to Real Definite Integrals. We shall illustrate, by examples, how the theory of residues

can be applied to determine integrals of real functions for special intervals.

Ex. 1. Consider the integral $\int e^{ix}dx/(x-i\beta)$. Let β be positive, and let the path be the real axis from $-\rho$ to ρ and a semicircle on this as diameter.

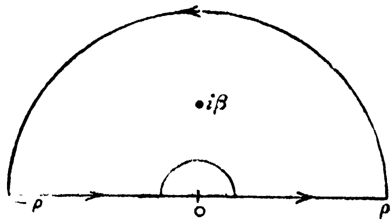


Fig. 53.

The integral along the base is $\int_{-\rho}^{\rho} e^{i\xi}d\xi/(\xi-i\beta)$; and along the semicircle it is $\int_0^{\pi} e^{ix}id\theta/(1-i\beta/x)$, where $x=\rho \operatorname{cis}\theta$.

When ρ tends to infinity this second integral tends to the limit 0, for $e^{ix}=e^{i\rho\cos\theta}e^{-\rho\sin\theta}$, and while the first factor is 1 in absolute value, the second tends to zero when $0<\theta<\pi$.

Also by the theory of § 119 the integral round the closed path is $2\pi i$. (the sum of the residues inside). There is only one residue, namely when $x=i\beta$, and this is $e^{-\beta}$; thus the integral round the path is $2\pi i e^{-\beta}$; and we have

$$\int_{-\infty}^{\infty} e^{i\xi}d\xi/(\xi-i\beta)=2\pi i e^{-\beta},$$

or
$$\int_{-\infty}^{\infty}\frac{\operatorname{cis}\xi d\xi\,(\xi+i\beta)}{\xi^2+\beta^2}=2\pi i e^{-\beta},$$

whence
$$\int_{-\infty}^{\infty}\frac{\beta\cos\xi+\xi\sin\xi}{\xi^2+\beta^2}d\xi=2\pi e^{-\beta}.$$

If we replace the positive number β by the negative number $-\beta$, then we have no residue inside the path, and therefore

$$\int_{-\infty}^{\infty}\frac{-\beta\cos\xi+\xi\sin\xi}{\xi^2+\beta^2}d\xi=0.$$

Hence
$$\beta\int_{-\infty}^{\infty}\frac{\cos\xi}{\xi^2+\beta^2}d\xi=\int_{-\infty}^{\infty}\frac{\xi\sin\xi}{\xi^2+\beta^2}d\xi=\pi e^{-\beta},$$

or
$$\int_{-\infty}^{\infty}\frac{\cos\xi}{\xi^2+\beta^2}d\xi=\pi e^{-\beta}/\beta.$$

Ex. 2. The equation $\int_{-\infty}^{\infty} \frac{\xi \sin \xi}{\xi^2 + \beta^2} d\xi = \pi e^{-\beta}$ is proved only for positive values of β; it suggests that when $\beta = 0$ we shall have $\int_{-\infty}^{\infty} \frac{\sin \xi}{\xi} d\xi = \pi$. But an integral found on the supposition that $i\beta$ is within the path must not be assumed to hold when $i\beta$ is on the path. When $\beta = 0$ we alter the path by describing a small semicircle about 0, say of radius ϵ as in fig. 53, and replacing the interval ($-\epsilon$ to ϵ) of the real axis by this semicircle. Then the integral is zero since there are no residues inside, and it reduces to

$$\int_{-\infty}^{-\epsilon} e^{i\xi} d\xi/\xi + \int_{\pi}^{0} e^{i\epsilon \operatorname{cis} \theta} i d\theta + \int_{\epsilon}^{\infty} e^{i\xi} d\xi/\xi.$$

The second integral tends to the limit $\int_{\pi}^{0} i d\theta$ or $-\pi i$ when ϵ tends to zero.

Hence $$\int_{-\infty}^{0} e^{i\xi} d\xi/\xi + \int_{0}^{\infty} e^{i\xi} d\xi/\xi = \pi i,$$

or $$\int_{-\infty}^{\infty} \sin \xi d\xi/\xi = \pi.$$

Ex. 3. The function $e^{ax}/(1 + e^x)$, where a is real, has poles at $x = (2n + 1)\pi i$. Integrate along the rectangle whose horizontal and vertical sides are given by $\eta = 0$, 2π, and $\xi = \pm \rho$.

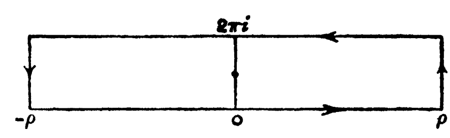


Fig. 54.

There is but one residue inside this rectangle, namely when $x = \pi i$; and this residue is $-e^{a\pi i}$.

The integrals along the vertical sides are

$$\int_{0}^{2\pi} \frac{e^{a(\rho+i\eta)} i d\eta}{1 + e^{\rho+i\eta}} \text{ and } \int_{2\pi}^{0} \frac{e^{a(-\rho+i\eta)} i d\eta}{1 + e^{-\rho+i\eta}}.$$

The limit of the first when ρ tends to ∞ will be zero if $a < 1$;

the limit of the second will be zero if $\alpha > 0$. Let then $0 < \alpha < 1$; there remain the integrals along the sides, that is,

$$\int_{-\infty}^{\infty} \frac{e^{\alpha\xi}}{1+e^{\xi}}\,d\xi + \int_{\infty}^{-\infty} \frac{e^{\alpha(2\pi i+\xi)}}{1+e^{\xi}}\,d\xi = -2\pi i e^{\alpha\pi i},$$

or
$$(1-e^{2\alpha\pi i})\int_{-\infty}^{\infty} \frac{e^{\alpha\xi}}{1+e^{\xi}}\,d\xi = -2\pi i e^{\alpha\pi i},$$

or
$$\int_{-\infty}^{\infty} \frac{e^{\alpha\xi}}{1+e^{\xi}}\,d\xi = \frac{2\pi i}{e^{\alpha\pi i}-e^{-\alpha\pi i}} = \pi/\sin\alpha\pi.$$

Ex. 4. If in the preceding example we replace α by $\alpha+i\beta$ or α, the argument is not altered provided $0 < \alpha < 1$. Hence

$$\int_{-\infty}^{\infty} \frac{e^{\alpha\xi}\operatorname{cis}\beta\xi}{1+e^{\xi}}\,d\xi = \pi/\sin(\alpha+i\beta)\pi,$$

or
$$\int_{-\infty}^{\infty} \frac{e^{(\alpha-\frac{1}{2})\xi}\operatorname{cis}\beta\xi}{2\cosh\xi/2}\,d\xi = \pi\frac{\sin(\alpha-i\beta)\pi}{\sin(\alpha+i\beta)\pi\sin(\alpha-i\beta)\pi}$$
$$= 2\pi\frac{\sin\alpha\pi\cosh\beta\pi - i\cos\alpha\pi\sinh\beta\pi}{\cosh 2\beta\pi - \cos 2\alpha\pi}$$

Hence, writing $\alpha-\frac{1}{2}=\mu$, so that $-1/2<\mu<1/2$, and equating real and imaginary parts, we get

$$\int_{-\infty}^{\infty} \frac{e^{\mu\xi}\cos\beta\xi}{2\cosh\xi/2}\,d\xi = 2\pi\frac{\cos\mu\pi\cosh\beta\pi}{\cosh 2\beta\pi+\cos 2\mu\pi},$$

or, changing the sign of μ and adding,

$$\int_{0}^{\infty} \frac{\cosh\mu\xi}{\cosh\xi/2}\cos\beta\xi\,d\xi = 2\pi\frac{\cos\mu\pi\cosh\beta\pi}{\cosh 2\beta\pi+\cos 2\mu\pi};$$

and
$$\int_{-\infty}^{\infty} \frac{e^{\mu\xi}\sin\beta\xi}{2\cosh\xi/2}\,d\xi = 2\pi\frac{\sin\mu\pi\sinh\beta\pi}{\cosh 2\beta\pi+\cos 2\mu\pi},$$

or
$$\int_{0}^{\infty} \frac{\sinh\mu\xi}{\cosh\xi/2}\sin\beta\xi\,d\xi = 2\pi\frac{\sin\mu\pi\sinh\beta\pi}{\cosh 2\beta\pi+\cos 2\mu\pi}.$$

Ex. 5. If in Ex. 3 two sides of the rectangle are $\eta=0$ and $\eta=\pi$, then the pole $x=\pi i$ is on the rectangle.

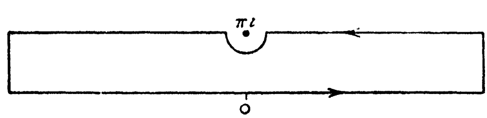


Fig. 55.

Avoiding the pole by a small semicircle of radius ϵ, and

observing that the integrals along the vertical sides tend to 0 when ρ tends to ∞ and $0 < \alpha < 1$, we have

$$\int_{-\infty}^{\infty} \frac{e^{\alpha\xi}}{1+e^{\xi}}\,d\xi + \int_{\infty}^{\epsilon} \frac{e^{\alpha(\xi+\pi i)}}{1-e^{\xi}}\,d\xi + \int_{-\epsilon}^{-\infty} \frac{e^{\alpha(\xi+\pi i)}}{1-e^{\xi}}\,d\xi + \int \frac{e^{ax}}{1-e^{x}}\,dx = 0,$$

the last term being the integral along the semicircle. This last integral is not itself half the integral round the whole circle, but its limit when ϵ tends to 0 is half the integral round that circle and is therefore equal to $+\pi i e^{a\pi i}$.

Hence understanding by $\int_{-\infty}^{\infty} \frac{e^{a\xi}}{1-e^{\xi}}\,d\xi$ the limit when ϵ tends to zero of

$$\left(\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}\right) \frac{e^{a\xi}}{1-e^{\xi}}\,d\xi,$$

we have
$$e^{a\pi i}\int_{-\infty}^{\infty} \frac{e^{a\xi}}{1-e^{\xi}}\,d\xi = \pi i e^{a\pi i} + \int_{-\infty}^{\infty} \frac{e^{a\xi}}{1+e^{\xi}}\,d\xi\,;$$

from which, equating imaginary parts, we deduce

$$\int_{-\infty}^{\infty} \frac{e^{a\xi}}{1-e^{\xi}}\,d\xi = \pi \cot a\pi.$$

By equating real parts we obtain again the result of Ex. 3, so that the rectangle of this example gives both integrals.

Ex. 6. When we replace α by $a = \alpha + i\beta$ we have still $\int_{-\infty}^{\infty} \frac{e^{a\xi}}{1-e^{\xi}}\,d\xi = \pi \cot a\pi$; for we have still $\int_{-\infty}^{\infty} \frac{e^{a\xi}}{1+e^{\xi}}\,d\xi = \frac{\pi}{\sin a\pi}$; and as this value led to $\pi \cot a\pi$ when a was real, it must do so when a is complex.

Hence, proceeding as in Ex. 4, we have

$$\int_{-\infty}^{\infty} \frac{e^{(\alpha-\frac{1}{2})\xi} \operatorname{cis} \beta\xi}{2\sinh \xi/2}\,d\xi = -\pi \cot(\alpha + i\beta)\,\pi,$$

or
$$\int_{-\infty}^{\infty} \frac{e^{\mu\xi} \operatorname{cis} \beta\xi}{2\sinh \xi/2}\,d\xi = \pi \tan(\mu + i\beta)\,\pi.$$

Hence
$$\int_{-\infty}^{\infty} \frac{e^{\mu\xi} \cos \beta\xi}{\sinh \xi/2}\,d\xi = \pi \tan(\mu + i\beta)\,\pi + \pi \tan(\mu - i\beta)\,\pi$$

$$= \frac{2\pi \sin 2\mu\pi}{\cos 2\mu\pi + \cosh 2\beta\pi};$$

whence, changing the sign of μ and subtracting,

$$\int_0^\infty \frac{\sinh \mu\xi}{\sinh \xi/2} \cos \beta\xi d\xi = \frac{\pi \sin 2\mu\pi}{\cos 2\mu\pi + \cosh 2\beta\pi}.$$

Similarly we have

$$i\int_{-\infty}^\infty \frac{e^{\mu\xi} \sin \beta\xi}{\sinh \xi/2} d\xi = \pi \tan (\mu + i\beta)\pi - \pi \tan (\mu - i\beta)\pi$$

$$= \frac{2\pi i \sinh 2\beta\pi}{\cos 2\mu\pi + \cosh 2\beta\pi},$$

whence $$\int_0^\infty \frac{\cosh \mu\xi}{\sinh \xi/2} \sin \beta\xi d\xi = \frac{\pi \sinh 2\beta\pi}{\cos 2\mu\pi + \cosh 2\beta\pi}.$$

The integrals in Examples 4 and 6 are required in discussing Fourier's Integral (see Byerly's *Fourier's Series and Spherical Harmonics*).

Ex. 7. Finally let us take $\int e^{-x^2}dx$ over the rectangle whose horizontal and vertical sides are $\eta = 0$, β and $\xi = \pm \alpha$.

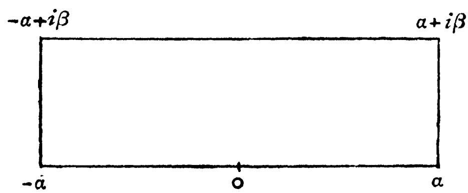


Fig. 56.

Within the rectangle e^{-x^2} has no singular points. Hence the integral is zero. Along two sides it becomes $\int_{-\alpha}^{\alpha} e^{-\xi^2}d\xi$ and $\int_{\alpha}^{-\alpha} e^{-(\xi+i\beta)^2}d\xi$; along the others it is $\int_0^\beta e^{-(\alpha+i\eta)^2}id\eta$ and $\int_\beta^0 e^{-(-\alpha+i\eta)^2}id\eta$. These last two tend to zero when α tends to infinity. Hence

$$\int_{-\infty}^\infty e^{-\xi^2}d\xi + \int_\infty^{-\infty} e^{-\xi^2-2i\beta\xi+\beta^2}d\xi = 0,$$

or $$e^{\beta^2}\int_{-\infty}^\infty e^{-\xi^2} \operatorname{cis}(-2\beta\xi) = \int_{-\infty}^\infty e^{-\xi^2}d\xi;$$

and the integral on the right is known to be $\pi^{1/2}$. Hence

$$\int_{-\infty}^{\infty} e^{-\xi^2} \cos 2\beta\xi d\xi = \pi^{1/2} e^{-\beta^2}.$$

Lest it should be supposed that the method of evaluating definite integrals, as described above, is in all cases the most suitable, we remark that it is difficult to determine $\int_{-\infty}^{\infty} e^{-\xi^2} d\xi$ itself by this method.

CHAPTER XVII.

LAURENT'S THEOREM AND THE THETA FUNCTIONS.

122. Laurent's Theorem. Let $fx = 1/(x-a)(x-b)$, where $0 < |a| < |b|$. The function has two simple infinities, namely $x=a$ and $x=b$. About the origin there is an element Px whose domain extends to a, and about ∞ there is an element $P(1/x)$ which converges at all points exterior to $(|b|)$. But there is also a series $Px + Q(1/x)$ which converges when $|a| < |x| < |b|$. That is, we draw circles with origin as centre through the singular points a, b, thus dividing the plane into the region inside $(|a|)$, the circular ring bounded by $(|a|)$, $(|b|)$, and the region outside $(|b|)$. In each of these regions there is a series of the form $\sum\limits_{n=-\infty}^{\infty} a_n x^n$, only it happens that in the inside region the negative powers of the corresponding series are absent and in the outside region there are no positive powers.

In this simple example the series for the three regions can be written down at once:—

$$(1) \qquad \frac{1}{(x-a)(x-b)} = \frac{1}{(a-b)(x-a)} - \frac{1}{(a-b)(x-b)};$$

hence, when $|x| < |a|$, we have

$$fx = -\frac{1}{(a-b)}\sum_{n=0}^{\infty}\left(\frac{1}{a^{n+1}} - \frac{1}{b^{n+1}}\right)x^n,$$

since $\qquad 1/(x-a) = -\sum\limits_{n=0}^{\infty} x^n/a^{n+1}, \quad 1/(x-b) = -\sum\limits_{n=0}^{\infty} x^n/b^{n+1}.$

(2) When $|x|>|b|$ we must use expansions in a/x and b/x; thus

$$fx=\frac{1}{a-b}\sum_{n=-1}^{-\infty}\left(\frac{1}{a^{n+1}}-\frac{1}{b^{n+1}}\right)x^n.$$

(3) For the ring we must use

$$1/(x-a)=\sum_{n=0}^{\infty}a^n/x^{n+1}=\sum_{n=-1}^{-\infty}x^n/a^{n+1}$$

$$1/(x-b)=-\sum_{n=0}^{\infty}x^n/b^{n+1}.$$

Hence $$fx=\frac{1}{a-b}\left[\sum_{n=-1}^{-\infty}x^n/a^{n+1}+\sum_{n=0}^{\infty}x^n/b^{n+1}\right]$$

Laurent's theorem is a generalization of this example for a one-valued analytic function fx which is analytic over an open region bounded by two concentric circles (R) and (R'), where $R>R'$. Let C, C' be any two circles about 0 which have radii slightly less than R and greater than R' respectively.

Applying Cauchy's integral to the ring bounded by C and C', we have

$$fc=\frac{1}{2\pi i}\int_C\frac{fxdx}{x-c}-\frac{1}{2\pi i}\int_{C'}\frac{fxdx}{x-c},$$

where c is a point within the ring and the circles are described positively with respect to the regions interior to them. For points x of C and points x of C' we use respectively

$$1/(x-c)=\quad 1/x+c/x^2+\ldots+c^n/x^{n+1}+c^{n+1}/x^{n+1}(x-c),$$
$$1/(x-c)=-1/c-x/c^2-\ldots-x^n/c^{n+1}-x^{n+1}/c^{n+1}(x-c).$$

Multiply the two series by fx; the resulting series converge uniformly and can be integrated term by term. Hence

$$fc=a_0+a_1c+a_2c^2+\ldots+a_nc^n+\ldots$$
$$+a_{-1}/c+a_{-2}/c^2+\ldots+a_{-n}/c^n+\ldots,$$

where $a_n=\frac{1}{2\pi i}\int_C fxdx/x^{n+1}$, $a_{-n}=\frac{1}{2\pi i}\int_{C'}fxdx\,.\,x^{n-1}$. If we wish to have a single formula for a_n and a_{-n} we can replace C, C' in the corresponding integrals by a concentric circle whose radius lies between those of C and C'. Thus replacing c by x, we have

for points x within the ring bounded by the circles (R), (R'), the expansion

$$\begin{aligned} fx &= a_0 + a_1 x + a_2 x^2 + \dots \text{ to infinity} \\ &\quad + a_{-1}x^{-1} + a_{-2}x^{-2} + \dots \text{ to infinity} \\ &= \sum_{n=-\infty}^{+\infty} a_n x^n. \end{aligned}$$

This is *Laurent's theorem*; and the associated series was called a *Laurent series* (§ 75). Of course if the centre of the ring had been a not 0, the Laurent series would have proceeded according to positive and negative powers of $x-a$.

Let R expand and R' shrink until the circles (R), (R') pass through singular points of fx. Then the first part of the Laurent series is a power series Px which has R for its radius of convergence, and the second part converges outside but not within (R').

If we add any two convergent series Px, $P(1/x)$, whose domains are the points within (R) and outside (R'), the special cases that can arise are (1) the two domains overlap, (2) they meet along a circle, (3) they do not meet. In case (3) the expression $Px+P(1/x)$ is meaningless, and in case (2) the points, if any, at which both series converge are confined to the circle of meeting.

To illustrate the possibility of a Laurent series being of the form (2), consider the series that arises from

$$\operatorname{Log}(1+x)=\frac{x}{1}-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\dots, \text{ where } |x| \leqq 1,\ x \neq -1,$$

$$\operatorname{Log}\left(1+\frac{1}{x}\right)=\frac{1}{x}-\frac{1}{2x^2}+\frac{1}{3x^3}-\frac{1}{4x^4}+\dots, \text{ where } |x| \geqq 1,\ x \neq -1.$$

By subtraction we have for $|x|=1$, $x=-1$ excepted,

$$\begin{aligned} \operatorname{Log} x &= \left(x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\dots\right)-\left(\frac{1}{x}-\frac{1}{2x^2}+\frac{1}{3x^3}-\frac{1}{4x^4}+\dots\right) \\ &= \frac{x-\frac{1}{x}}{1}-\frac{x^2-\frac{1}{x^2}}{2}+\frac{x^3-\frac{1}{x^3}}{3}-\frac{x^4-\frac{1}{x^4}}{4}+\dots. \end{aligned}$$

By writing $x=e^{i\eta}$, we get again the formula of § 96, namely

$$\frac{\eta}{2}=\frac{\sin\eta}{1}-\frac{\sin 2\eta}{2}+\frac{\sin 3\eta}{3}-\frac{\sin 4\eta}{4}+\dots,$$

where η is real and $-\pi<\eta<+\pi$.

123. Isolated Singularities of One-valued Functions. It happens frequently that a function is known to be analytic

about all points of a region except possibly certain isolated points; it is desirable to have some means of inferring the behaviour of a one-valued analytic function fx *about* such points from the behaviour of the function *near* them.

Laurent's theorem gives important information as we shall show.

Let fx be analytic about all points, other than c, of the closed region (c, ρ); and let it be unknown whether c is an ordinary or a singular point or a point at which fx is undefined. Whatever be the character of c we can find a Laurent series for fx in the ring bounded externally by (c, ρ) and internally by an arbitrarily small circle of centre c; furthermore because the inner circle can shrink indefinitely the negative powers in the series define a function $G\left(\frac{1}{x-c}\right)$. Only three cases can arise:—

(1) The Laurent series may contain no negative powers of $x-c$; then

$$fx = P(x-c)$$

near $x=c$.

(2) It may contain negative powers in finite number; then

$$fx = (x-c)^{-n} P_0(x-c)$$

near $x=c$.

(3) It may be of the form

$$fx = G(1/(x-c)) + P(x-c)$$

near $x=c$, Gx denoting a transcendental integral function.

Cases (1), (2), (3) are characterized by the following distinctive properties:—in (1) we have $\lim\limits_{x=c} fx =$ the constant term of $P(x-c)$, in (2) we have $\lim\limits_{x=c} fx = \infty$, and in (3) there is no one definite limit for fx when x approaches c.

In ease (1) we know that $|fx|$ remains constantly below some finite positive number γ; and now fx at c is defined or altered in value so as to make $fc = \lim\limits_{x=c} fx$.

In case (2) where $\lim\limits_{x=c} fx = \infty$ whatever be the path of approach

to c, it is natural to define or alter the value of fx at c so as to make it ∞.

We shall now prove Weierstrass's theorem (§ 104) that *near an essential singularity which is isolated as regards other essential singularities a one-valued analytic function fx approaches as near as we please to an arbitrarily given value f_0.* This can be proved by considering $1/(fx-f_0)$ near its essential singularity c. Either $1/(fx-f_0)$ has an infinity of non-essential singularities which have c as a limit-point,—in which case fx takes the *exact* value f_0 at points as near to c as we please,—or else c is an isolated singularity of $1/(fx-f_0)$. In this latter case there is a Laurent series for points near $x=c$ of the form

$$\frac{1}{fx-f_0}=G(1/(x-c))+P(x-c).$$

But near $x=c$ we can make $G(1/(x-c))$ exceed in absolute value any assigned positive number γ however large (§ 93); hence we can make fx approach as near as we please to f_0 even though it may not attain f_0. This proves the theorem.

Let us examine, in the light of this theorem, the behaviour for case (3) of fx near $x=c$. We cannot speak now of $\lim_{x=c} fx$ without specifying the path of approach to c; in fact along some paths there would not be any limiting value. For this reason it is useless to define fc as $\lim_{x=c} fx$; hence fx is left undefined at c (§ 95).

Observe that in all three cases the point $x=c$, if singular at all, has been supposed to be an isolated singular point. For example when $c=0$ the function $\csc 1/x$ has an essential singularity at $x=c$, but $\csc 1/x$ cannot be put in the form

$$G(1/x)+Px,$$

and the preceding work does not apply. To avoid misconceptions notice that whereas here there is no Laurent series for the ring formed from a circle (R) by the removal of $x=0$, there are Laurent series for each of the annular regions extending from ∞ to $1/\pi$, $1/\pi$ to $1/2\pi$, $1/2\pi$ to $1/3\pi$, ..., these regions becoming infinitely congested near $x=0$.

Let us now consider whether the Laurent series for fx in a given ring bounded by (c, R_2), (c, R_1), where $R_2 < R_1$, is or is not unique. Suppose that

$$fx = \sum_{n=-\infty}^{\infty} a_n (x-c)^n.$$

Divide both sides by $(x-c)^{n+1}$ and integrate term by term over any circle (c, R) which lies within the ring; this is permissible because $\sum_{n=-\infty}^{-1} a_n (x-c)^n$ converges uniformly along (c, R) because $R > R_2$, and $\sum_{n=0}^{\infty} a_n (x-c)^n$ converges uniformly along (c, R) because $R < R_1$. Hence

$$\int_{(c,R)} fx\,(x-c)^{-n-1}dx = 2\pi i a_n;$$

thus a_n has the same value as the coefficient for $(x-c)^n$ when we use Laurent's method, and the Laurent series for the given ring proves to be the only series of the form $\sum_{n=-\infty}^{\infty} a_n (x-c)^n$ that will represent fx in that ring.

124. Fourier's Series. From a Laurent series

$$fx = \sum_{n=-\infty}^{\infty} a_n x^n$$

we can deduce a Fourier series by writing $x = e^{iy}$, and then replacing e^{niy} by $\cos ny + i \sin ny$. Let us examine this more closely.

Let the radii of the ring be e^{α}, e^{β} where α, β are real and $|\alpha| < |\beta|$; and let $y = \xi' + i\eta'$. When $x = e^{\alpha}$, e^{β} we can take $\eta' = -\alpha$, $-\beta$; also when $\operatorname{Am} x$ passes from $-\pi$ to $+\pi$, ξ' passes from $-\pi$ to $+\pi$. The ring, when supposed cut along the real axis from $-e^{\alpha}$ to $-e^{\beta}$, maps into the rectangle given by the equations

$$\xi' = -\pi,\ \xi' = +\pi,\ \eta' = -\beta,\ \eta' = -\alpha.$$

Within this rectangle we have for the one-valued analytic function ϕy of y, given by

$$\phi y = f(e^{yi}),$$

the expansion

$$\sum_{n=-\infty}^{\infty} a_n e^{nyi}, \text{ or } \sum_{n=-\infty}^{\infty} a_n (\cos ny + i \sin ny).$$

By combining terms that arise from opposite n's, we have

$$\phi y = a_0 + \sum_{n=1}^{\infty} (b_n \cos ny + c_n \sin ny),$$

a function which is analytic over the rectangle.

Conversely $x = e^{yi}$ converts the y-rectangle into the cut x-ring, *the cut going from* $-e^{\alpha}$ *to* $-e^{\beta}$, and any analytic function ϕy which is analytic over the rectangular region changes into an analytic function fx which is analytic over the cut annular region. If we are to replace the cut by the uncut ring, it is necessary that fx shall take the same value at a point of the cut from $-e^{\alpha}$ to $-e^{\beta}$ whether the approach to this point be from above or below, and be analytic about all such points. This equality of values is secured immediately by making ϕy a periodic function with the period 2π.

When ϕy is periodic with the period 2π we can translate the rectangle through strokes $\pm 2n\pi$ and thus form a band of the y-plane bounded by the two parallel lines $\eta' = -\alpha$, $\eta' = -\beta$, and see at once that ϕy is analytic over the band. The previous expansion for ϕy in terms of sines and cosines holds for the whole of this band; or again we may say that ϕy is expansible in the form

$$\sum_{n=-\infty}^{\infty} a_n e^{nyi}.$$

The value of a_n is determined from the Laurent series, and is

$$\frac{1}{2\pi i}\int z^{-(n+1)} fz\, dz, \quad = \frac{1}{2\pi}\int e^{-nzi} \phi z\, dz,$$

the integration extending over any x-circle about 0 whose radius lies between e^{α} and e^{β}.

By replacing e^{nyi}, etc. by $\cos ny + i \sin ny$, etc. and combining the terms in n and $-n$, we get

$$\frac{1}{2\pi} (\cos ny + i \sin ny) \int \phi z (\cos nz - i \sin nz)\, dz$$

$$+ \frac{1}{2\pi} (\cos ny - i \sin ny) \int \phi z (\cos nz + i \sin nz)\, dz,$$

that is,

$$\frac{1}{\pi}\int \phi z\,(\cos ny \cos nz + \sin ny \sin nz)\,dz,$$

or
$$\frac{1}{\pi}\int \phi z \cos n\,(y-z)\,dz.$$

When $n=0$ we have the single term $\frac{1}{2\pi}\int \phi z dz$. Combining these results

$$\phi y = \frac{1}{\pi}\int\left[1 + 2\sum_{n=1}^{\infty} \cos n\,(y-z)\right]\phi z\,dz.$$

This is the ordinary form for Fourier's series; it should be remarked however that it has been established under conditions that differ from those that usually occur; the theorem in the present form holds for complex as well as for real values, but this gain in generality is accompanied by a loss in another direction, due to the restrictions on ϕy.

125. The Partition-function. We can write down by the factor theorem, ch. XV., a function which is transcendental and integral, and is 0^1 at the points

$$1,\ 1/r,\ 1/r^2, \ldots\ldots,\ 1/r^n, \ldots\ldots,$$

when $\mathrm{L}\ 1/r^n$ is ∞, that is, when $|r| < 1$. The simplest function with these properties is

$$Gx = (1-x)(1-rx)(1-r^2x)\ldots(1-r^nx)\ldots,$$

the product being convergent for any x because of the convergence of

$$1 + r + r^2 + \ldots.$$

This function,—which is fundamental in the theory of partitions of integers*,—can be written as a power series

$$1 + a_1x + a_2x^2 + \ldots;$$

and the coefficients a_n are readily determined by the equation

$$(1-x)\,G(rx) = Gx,$$

whence
$$a_nr^n - a_{n-1}r^{n-1} = a_n,$$

* See Chrystal's *Algebra*, vol. ii. p. 528.

or
$$a_n = a_{n-1} r^{n-1}/(r^n - 1),$$
$$= \frac{r^{n-1+n-2+\ldots+1}}{(r^n - 1)(r^{n-1} \cdot\cdot 1)\ldots(r-1)},$$
$$= \frac{r^{n(n-1)/2}}{(r^n - 1)(r^{n-1} - 1)\ldots(r-1)}.$$

Let us complete the series of zeros by the points $r, r^2, \ldots, r^n, \ldots$ whose limit is 0.

A function with these zeros will be $G(r/x)$ where Gx is the preceding function. The product $fx = GxG(r/x)$ will be one-valued and analytic about all points except 0 and ∞. By Laurent's theorem it can be developed in the form $\sum_{-\infty}^{\infty} b_n x^n$ To determine the coefficients b_n we have
$$f(rx) = G(rx)\, G(1/x);$$
or, since
$$G(rx) = \frac{Gx}{1-x},$$
and
$$G(r/x) = \frac{G(1/x)}{1 - 1/x},$$
therefore
$$xf(rx) = -\, GxG(r/x) = -fx.$$

Hence
$$b_{n-1} r^{n-1} = -\, b_n,$$
so that we get when n is positive,
$$b_n = (-)^n\, b_0 r^{n-1+n-2+\ldots+1} = (-)^n\, b_0 r^{n(n-1)/2};$$
and when n is negative and $= 1 - m$,
$$b_{-m} = -\, b_{1-m} r^m,$$
$$= (-)^m\, b_0 r^{m+m-1+\ldots+1},$$
$$= (-)^m\, b_0 r^{m(m+1)/2},$$
$$= (-)^{-m}\, b_0 r^{-m(-m-1)/2}.$$

Thus, whatever integer n may be, we have
$$b_n = (-)^n\, b_0 r^{n(n-1)/2},$$
and the Laurent series for fx is
$$fx = b_0 \sum_{-\infty}^{\infty} (-)^n\, r^{n(n-1)/2} x^n.$$

To determine the constant b_0, we can proceed as follows. We have
$$fx = Gx\, G(r/x),$$
$$= (1 + a_1 x + a_2 x^2 + \ldots)(1 + a_1 r/x + a_2 r^2/x^2 + \ldots),$$

where $$a_n = (-)^n r^{n(n-1)/2} / \prod_1^n (1 - r^p).$$

Hence, multiplying out and comparing with $\Sigma b_n x^n$,

$$b_n = a_n + a_1 a_{n+1} r + a_2 a_{n+2} r^2 + \ldots = \sum_{m=0}^{\infty} a_m a_{n+m} r^m,$$

where $a_0 = 1$.

Therefore, replacing a_m and b_n by the values found, we have, on dividing by $(-)^n r^{n(n-1)/2}$,

$$b_0 = \sum_{m=0}^{\infty} \frac{r^{m+m(m-1)/2+(n+m)(n+m-1)/2-n(n-1)/2}}{\prod_1^m (1-r^p) \prod_1^{n+m} (1-r^p)},$$

$$= \sum_{m=0}^{\infty} \frac{r^{m(m+n)}}{\prod_1^m (1-r^p) \prod_1^{n+m} (1-r^p)},$$

whatever positive integer n may be.

Omitting the first term the series can be expressed as the product of a convergent series and r^n; hence when n tends to ∞ the sum of the terms after the first tends to zero on account of the factor r^n, and we obtain for the constant b_0 the expression

$$1 / \prod_1^{\infty} (1 - r^p).$$

To sum up, if $|r| < 0$ and if

$$Gx = (1-x)(1-rx)(1-r^2x)\ldots = \prod_0^{\infty} (1 - r^p x),$$

then $$GxG(r/x) = b_0 \sum_{-\infty}^{\infty} (-)^n r^{n(n-1)/2} x^n,$$

where $$b_0 = 1 / \prod_1^{\infty} (1 - r^p) = 1/Gr.$$

Ex. 1. By putting $x = -1$, prove that

$$1 + r + r^3 + \ldots + r^{n(n-1)/2} + \ldots = \prod_1^{\infty} (1 + r^p)(1 - r^{2p}).$$

Verify this for the first few powers by actual multiplication of the product on the right-hand side.

Ex. 2. Prove that when $\sum_{-\infty}^{\infty} (-)^n r^{n(n-1)/2} x^n$ is divided by $1 - x$, the quotient is

$$1 - r(x + 1 + 1/x) + r^3(x^2 + x + 1 + 1/x + 1/x^2) - \ldots;$$

and hence by writing $x=1$ prove that

$$1-3r+5r^3-\ldots-(-)^n(2n-1)r^{n(n-1)/2}+\ldots=\prod_1^\infty(1-r^p)^3.$$

Ex. 3. Express as infinite products

$$x+x^9+x^{25}+x^{49}+\ldots,$$

and
$$x-3x^9+5x^{25}-7x^{49}+\ldots,$$
where
$$|x|<1.$$

126. The Theta Functions. It is convenient to replace r by q^2, so that the zeros of fx are now at the points q^{2m} where m is zero or any integer, positive or negative. We have then

$$fx=b_0\sum_{-\infty}^{\infty}q^{n^2}(-x/q)^n$$
$$=b_0+b_0\sum_1^\infty q^{n^2}[(-x/q)^n+(-x/q)^{-n}].$$

If then $-x/q=\exp 2\pi iv$, we have for fx the Fourier series

$$b_0+2b_0\sum_1^\infty q^{n^2}\cos 2n\pi v.$$

The ring for the Laurent series was the whole plane except $x=0$ and $x=\infty$; hence the rectangle for the Fourier series is the whole plane except ∞. That is, the Fourier series defines a transcendental integral function; it has a simple zero when $\exp 2\pi iv=-q^{2m-1}$, that is, if $q=\exp\pi i\omega$, when

$$\exp 2\pi iv=\exp[(2m-1)\pi i\omega+(2m'-1)\pi i],$$

or when
$$v=-(1+\omega)/2+m\omega+m',$$

where m and m' are any integers or zero.

Omitting the factor b_0 we have a function $\vartheta_3 v$, with the specified properties, fully defined for a finite v by the equation

$$\vartheta_3 v=1+2q\cos 2\pi v+2q^4\cos 4\pi v+2q^9\cos 6\pi v+\ldots,$$

it being understood that $|q|<1$.

From the definition of $\vartheta_3 v$,

$$\vartheta_3(v+1/2)=1-2q\cos 2\pi v+2q^4\cos 4\pi v-2q^9\cos 6\pi v+\ldots;$$

for this allied function we write $\vartheta_2 v$. It has simple zeros at the points $-\omega/2+m\omega+m'$.

The functions $\vartheta_3 v$ and $\vartheta_2 v$ are on the face of them singly periodic; they have both the period 1. Therefore also $\vartheta_3 v/\vartheta_2 v$ has the period 1. We shall prove that it has also the period 2ω.

We have $$fx = b_0 \vartheta_3 v,$$
where $$x = -q \exp 2\pi i v.$$

When we add $\omega/2$ to v, we multiply x by $\exp \pi i \omega$ or q.

Hence $$f(qx) = b_0 \vartheta_3 (v + \omega/2),$$
and $$f(q^2 x) = b_0 \vartheta_3 (v + \omega).$$
Now $$xf(q^2 x) = -fx;$$
hence $$-\exp 2\pi i (v + \omega/2)\, \vartheta_3 (v + \omega) = \vartheta_3 v.$$
Changing v into $v + \frac{1}{2}$, we deduce
$$\exp 2\pi i (v + \omega/2)\, \vartheta_2 (v + \omega) = \vartheta_2 v.$$
Hence $$\vartheta_3 (v + \omega)/\vartheta_2 (v + \omega) = -\vartheta_3 v/\vartheta_2 v;$$
that is, $\vartheta_3 v/\vartheta_2 v$ changes sign when we increase v by ω, and therefore repeats its value when we increase v by 2ω.

Hence $\vartheta_3 v/\vartheta_2 v$ is a doubly periodic transcendental fractional function, having the two periods 1 and 2ω.

If $\omega = \alpha + i\beta$, then $q = \exp \pi i \alpha \exp - \pi\beta$, and $|q| = \exp - \pi\beta$. The condition $|q| < 1$ requires, then, that β be positive. Thus the period 2ω is complex or imaginary, with a positive coefficient of i.

We have so far defined only two ϑ-functions, while the notation suggests that there are others.

Returning to the function fx, or $b_0 \Sigma (-)^n q^{n^2-n} x^n$, we observe that for each exponent $n^2 - n$ of q there are two exponents of x whose sum is 1. Thus
$$fx = b_0 \sum_1^\infty (-)^n q^{n^2-n} (x^n - x^{1-n}),$$
or, replacing x by z^2,
$$fx = b_0 z \sum_1^\infty (-)^n q^{n^2-n} (z^{2n-1} - z^{1-2n}).$$

If then $z = \exp \pi i v$ we have $z^{2n-1} - z^{1-2n} = 2i \sin (2n - 1) \pi v$,

and $$f(z^2) = b_0 \frac{z}{i} [2 \sin \pi v - 2q^2 \sin 3\pi v + 2q^6 \sin 5\pi v - \ldots]$$
$$= b_0 \frac{z}{i} q^{-1/4} [2q^{1/4} \sin \pi v - 2q^{9/4} \sin 3\pi v + 2q^{25/4} \sin 5\pi v - \ldots].$$

The series in square brackets will be* denoted by ϑv. If we write $v + 1/2$ for v (equivalent to iz for z), we obtain

$$2q^{1/4}\cos\pi v + 2q^{9/4}\cos 3\pi v + 2q^{25/4}\cos 5\pi v + \dots,$$

a series which will be denoted by $\vartheta_1 v$.

For convenience of reference we repeat the formulae for the four ϑ-functions: they are

$$\vartheta_3 v = 1 + 2q\cos 2\pi v + 2q^4\cos 4\pi v + 2q^9\cos 6\pi v + \dots,$$

$$\vartheta_2 v = 1 - 2q\cos 2\pi v + 2q^4\cos 4\pi v - 2q^9\cos 6\pi v + \dots,$$

$$\vartheta_1 v = 2q^{1/4}\cos\pi v + 2q^{9/4}\cos 3\pi v + 2q^{25/4}\cos 5\pi v + \dots,$$

$$\vartheta v \ = 2q^{1/4}\sin\pi v + 2q^{9/4}\sin 3\pi v + 2q^{25/4}\sin 5\pi v - \dots;$$

of these functions the first three are even, the fourth odd.

Detailed information on these functions will be found in works on elliptic functions. We now take up a more symmetric function with the same arrangement of zeros as the ϑ-functions.

* The reader must be warned that in the literature of the subject there is a diversity of notations. The notation adopted here is that used by Jordan in his *Cours d'Analyse*, vol. ii. (1894); it differs from that employed in our *Treatise on the Theory of Functions*, (1893).

CHAPTER XVIII.

FUNCTIONS ARISING FROM A NETWORK.

127. The Network. All the points of a plane which are given by the formula

$$w = 2m_1\omega_1 + 2m_2\omega_2,$$

where $2\omega_1$ and $2\omega_2$ are given constant complex numbers and m_1 and m_2 are any integers, positive, zero, or negative, are said to

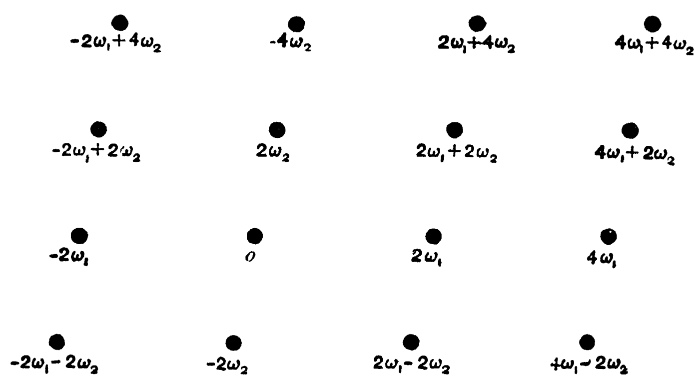


Fig. 57.

form a *network* (fig. 57). We suppose that the ratio ω_2/ω_1 is not real; thus avoiding the degenerate case in which all the points lie on a line.

We can call w a multiple of $2\omega_1$ and $2\omega_2$. Any two numbers which differ by a number w are said to be congruent one with the other. The sign $\equiv$ means 'is congruent with.'

Now we can select from among the w's pairs of numbers other than $2\omega_1$ and $2\omega_2$ which will equally give the same network; for example $2\omega_1$ and $2(\omega_1+\omega_2)$. For clearly $2m_1\omega_1 + 2m_2(\omega_1+\omega_2)$

gives the same points as $2m_1\omega_1 + 2m_2\omega_2$, though of course not in the same order. We call any pair of the numbers w, with which the same network can be built, a primitive pair. The condition that a selected pair, say $2m_1\omega_1 + 2m_2\omega_2$ and $2m'_1\omega_1 + 2m'_2\omega_2$, shall be primitive is that, if $\Delta = \begin{vmatrix} m_1, m'_1 \\ m_2, m'_2 \end{vmatrix}$, then $\Delta = \pm 1$.

For if $\quad m_1\omega_1 + m_2\omega_2 = \omega'_1, \quad m'_1\omega_1 + m'_2\omega_2 = \omega'_2,$

then $\quad \Delta\omega_1 = m'_2\omega'_1 - m_2\omega'_2, \quad \Delta\omega_2 = m_1\omega'_2 - m'_1\omega'_1,$

whence ω_1 and ω_2 are multiples of ω'_1 and ω'_2 when and only when $\Delta = \pm 1$. When this condition is satisfied, we can recover the original pair ω_1 and ω_2, and their multiples will equally be multiples of ω'_1 and ω'_2. Thus the network formed with ω'_1 and ω'_2 is the same as that formed with ω_1 and ω_2.

The geometric meaning is easily seen to be that if we complete the parallelogram whose adjacent sides are a primitive pair, the area is the same whatever primitive pair we take. The region enclosed by such a parallelogram of periods will be called a *cell*.

Returning to our given numbers $2\omega_1$ and $2\omega_2$, we suppose henceforth, unless the contrary is expressly stated, that the points 0, $2\omega_1$, $2(\omega_1 + \omega_2)$, $2\omega_2$, which are the corners of a cell, lie in positive order, that is, that a point which describes the parallelogram so that the enclosed region is on the left meets the points in the order named. If this were not so we should merely interchange ω_1 and ω_2. This is equivalent to saying that $\text{Am}\,\omega_2/\omega_1$ is positive, or again to saying that if $\omega_2/\omega_1 = \alpha + i\beta$, then $\beta > 0$. This convention is essential when we connect the functions of this chapter with the theta-functions. To ω_1 and ω_2 we add for symmetry a third number ω_3 such that

$$\omega_1 + \omega_2 + \omega_3 = 0,$$

and fig. 58 shows that $\text{Am}\,\omega_3/\omega_2$, $\text{Am}\,\omega_1/\omega_3$ are positive.

For a primitive pair ω'_1, ω'_2 we now require that $\text{Am}\,\omega'_2/\omega'_1$ shall be positive; hence $\Delta = +1$. The following are primitive pairs: $2\omega_1$ and $2\omega_2$ by hypothesis, $2\omega_3$ and $2\omega_1$ because here

$$\omega_3 = -\omega_1 - \omega_2,$$

$$\omega_1 = +\omega_1,$$

so that
$$\begin{vmatrix} m_1, & m'_1 \\ m_2, & m'_2 \end{vmatrix} = \begin{vmatrix} -1, & 1 \\ -1, & 0 \end{vmatrix} = 1,$$
and $2\omega_2$ and $2\omega_3$ from a similar calculation.

If u be any point of the plane, the points $u + 2m_1\omega_1 + 2m_2\omega_2$ form a network congruent with u.

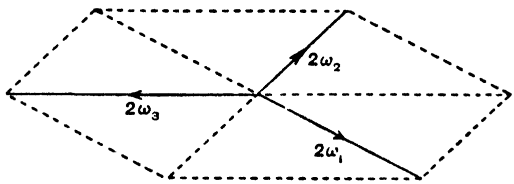


Fig. 58.

128. A Theorem on Convergence. We know that the series
$$1/1^2 + 1/2^2 + 1/3^2 + \ldots,$$
is convergent; its sum in fact is $\pi^2/6$.

It follows that if λ is an integer greater than 1, $\sum\limits_{-\infty}^{\infty}{}' 1/m^\lambda$ is convergent, since each term is diminished by increasing λ.

It is important to prove the analogous theorem for a network, that $\Sigma' 1/|w|^\lambda$ is convergent when the integer $\lambda > 2$; the summation being for all points $w = 2m_1\omega_1 + 2m_2\omega_2$, except the origin.

It is to be remarked that these convergence theorems are true for all real values of λ, > 1 in the first case, and > 2 in the second case, but we are concerned only with integer values.

The points w (fig. 57) lie on the sides of the following parallelograms with centres at the origin. The first parallelogram has for its corners $\pm 2\omega_1 \pm 2\omega_2$ and there are 8 points on it. Let δ be the least distance from the origin to this parallelogram; then for the 8 points
$$\Sigma 1/|w|^\lambda < 8/\delta^\lambda.$$

The second parallelogram has corners at $\pm 4\omega_1 \pm 4\omega_2$, and there are on it $8 \cdot 2$ points. Its least distance from the origin is 2δ, hence for these $8 \cdot 2$ points
$$\Sigma 1/|w|^\lambda < 8 \cdot 2/(2\delta)^\lambda.$$

The nth parallelogram has corners at $\pm 2n\omega_1 \pm 2n\omega_2$; there are $8n$ points on it, and its least distance from the origin is $n\delta$. Hence for these $8n$ points

$$\Sigma 1/|w|^\lambda < 8n/(n\delta)^\lambda.$$

Using the network in this way, we can convert $\Sigma' 1/|w|^\lambda$ into a single series and see that

$$\begin{aligned}\Sigma' 1/|w|^\lambda &< 8/\delta^\lambda + 8 \cdot 2/(2\delta)^\lambda + \ldots + 8n/(n\delta)^\lambda + \ldots, \\ &< 8(1 + 1/2^{\lambda-1} + \ldots + 1/n^{\lambda-1} + \ldots)/\delta^\lambda.\end{aligned}$$

This latter series is convergent if $\lambda - 1 > 1$, that is, if $\lambda > 2$. Hence the series $\Sigma' 1/|w|^\lambda$ is convergent if $\lambda > 2$.

Similar reasoning, when δ is replaced by the greatest distance from 0 to the first parallelogram, shows that $\Sigma' 1/|w|^\lambda$ is divergent if $\lambda = 2$.

129. The Functions σu, ζu, $\wp u$. We have found in $\sin u$ a simple example of a transcendental integral function of grade 1; we proceed to the construction of a transcendental integral function of grade 2.

Since $\Sigma' 1/|w|^2$ diverges and $\Sigma' 1/|w|^3$ converges, the transcendental integral function which is 0^1 at, and only at, the points w must be of the second grade (§ 111); the typical factor is therefore

$$(1 - u/w) \exp(u/w + u^2/2w^2),$$

and *the infinite product*

$$u\Pi' \{(1 - u/w) \exp(u/w + u^2/2w^2)\},$$

where the accent signifies the exclusion of $w = 0$, *is a transcendental integral function which answers the requirements.* This function is denoted by σu and called the *σ-function.* It is of capital importance in Weierstrass's exposition of elliptic functions.

The exponential, it will be remembered, was introduced into the primary factor to ensure convergence. In the present case

$$\begin{aligned}u\Pi'(1 - u/w) &= u\Pi' e^{\mathrm{Log}(1-u/w)} \\ &= u \exp\left\{-u\Sigma'\frac{1}{w} - \frac{u^2}{2}\Sigma'\frac{1}{w^2} - \ldots\right\};\end{aligned}$$

knowing that $\Sigma' 1/w$, $\Sigma' 1/w^2$ become divergent when w is replaced by $|w|$, we remove them by the introduction of the factors $\exp(u/w + u^2/2w^2)$, and leave only $\Sigma' 1/w^3$, $\Sigma' 1/w^4$, etc., all of which converge when w is replaced by $|w|$.

By § 111 the most general transcendental integral function with the same zeros as σu, and to the same order, is $e^{Gu}\sigma u$.

Taking (§ 113) the logarithms of both sides of

$$\sigma u = u\Pi'\left[(1 - u/w)\exp(u/w + u^2/2w^2)\right] \quad \ldots\ldots\ldots (1),$$

and then the derivate as to u we have, denoting the derivate of σu by $\sigma' u$,

$$\frac{\sigma' u}{\sigma u} = \frac{1}{u} + \Sigma'\left[\frac{1}{u - w} + \frac{1}{w} + \frac{u}{w^2}\right].$$

Here the product Π' has become a sum Σ' extended, equally with the product, over all values of w except the value $w = 0$. This is a function which is one-valued and finite at all finite points except the points w, at which it has simple infinities. It is denoted by ζu and called the *ζ-function*. Thus

$$\zeta u = \frac{1}{u} + \Sigma'\left[\frac{1}{u - w} + \frac{1}{w} + \frac{u}{w^2}\right] \quad \ldots\ldots\ldots\ldots\ldots (2).$$

The derivate of the function ζu is given by

$$\zeta' u = -\frac{1}{u^2} + \Sigma'\left[-\frac{1}{(u - w)^2} + \frac{1}{w^2}\right],$$

and this with changed sign is denoted by $\wp u$, pronounced p, u. This function is called the $\wp$-function. Thus

$$\wp u = \frac{1}{u^2} + \Sigma'\left[\frac{1}{(u - w)^2} - \frac{1}{w^2}\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots (3).$$

That the series (2) and (3) converge absolutely and uniformly in every closed region Γ of the u-plane which contains no point w, can be proved directly. Taking the series in (3), and noticing that the general term

$$\frac{1}{(u - w)^2} - \frac{1}{w^2}, = \frac{2u}{w^3}\,\frac{1 - \dfrac{u}{2w}}{\left(1 - \dfrac{u}{w}\right)^2},$$

has $2u/w^3$ for its asymptotic value when w tends to infinity, we see that the convergence of $\wp u$ must be absolute because $\Sigma' 2u/w^3$ converges absolutely. But the convergence is also uniform; for the absolute values of the points u of Γ have a finite upper limit A; hence, taking for the α's of § 73 the numbers $2A/|w|^3$, the series $\Sigma'\{1/(u - w)^2 - 1/w^2\}$ is seen to fulfil

the requirements of that article and therefore must converge uniformly.

Ex. Prove directly that the series for ζu, and the series $\Sigma\, 1/(u-w)^r$ where $r=3, 4, \ldots$, converge uniformly in the region Γ defined as above.

The formulæ

$$\sigma u = u\Pi' \{(1 - u/w) \exp (u/w + u^2/2w^2)\}$$
$$\zeta u = 1/u + \Sigma' \{1/(u-w) + 1/w + u/w^2\}$$
$$\wp u = 1/u^2 + \Sigma' \{1/(u-w)^2 - 1/w^2\}$$

show that σu, ζu are odd functions, $\wp u$ an even function, that the residues of ζu at its infinities are all equal to 1,—or, as it is often stated, ζu is infinite at $u=0$ like $1/u$ and at $u=w$ like $1/(u-w)$,—and that the infinities of $\wp u$ are all of order 2 with zero residues. Further they show that

$$\zeta' u = -\wp u,$$
$$\zeta u = D \log \sigma u,$$
$$\wp u = -D^2 \log \sigma u.$$

When σu, ζu, $\wp u$ are regarded as functions of three variables u, $2\omega_1$, $2\omega_2$, the functions become homogeneous of degrees 1, -1, -2. For example every factor in $\sigma u/u$ is homogeneous of degree 0, and therefore σu is homogeneous and of degree 1.

The derivate of $\wp u$ is

$$\wp' u = -\frac{2}{u^3} - \Sigma' \frac{2}{(u-w)^3} = -2\Sigma \frac{1}{(u-w)^3} \quad \ldots\ldots\ldots(4),$$

an odd function with a triple infinity at each point w and no other infinities.

We have derived ζu, $\wp u$, $\wp' u$ from σu, but in defining these functions we might have begun with any one of them. For example suppose that we seek first of all for a transcendental fractional function which shall be infinite like $-2/(u-w)^3$ at each of the points w and shall have no other singular points in the finite part of the plane. The series $-2\Sigma\, 1/(u-w)^3$ answers the requirements; for this series is uniformly convergent in the region Γ defined as above and it defines an analytic function which is analytic about every point a which is not equal to a w,

since Weierstrass's theorem of § 81 applies to the infinite series of power series for which the general term is

$$-2/(u-w)^3 = -2/(u-c-w+c)^3$$
$$= 2\left\{\frac{1}{(w-c)^3} + \frac{3u}{(w-c)^4} + \dots\right\}.$$

Defining $\wp' u$ by $-2\Sigma\, 1/(u-w)^3$, the value of $\wp u$ is found by integration, and from it in turn it is possible to deduce ζu and σu.

130. Series for $\wp u$, ζu, σu in powers of u. Let a closed region (ρ) contain no point w except 0; then if we expand $1/(u-w)$ in powers of u we have in (ρ)

$$\frac{1}{u-w} = -\frac{1}{w}\left(1-\frac{u}{w}\right)^{-1}$$
$$= -1/w - u/w^2 - u^2/w^3 - u^3/w^4 - \dots.$$

Hence $$\zeta u = 1/u - u^2\Sigma'\, 1/w^3 - u^3\Sigma'\, 1/w^4 - \dots.$$

The absolutely convergent w-series $\Sigma'\, 1/w^{2n+1}$, where $n = 1, 2, 3, \dots$, is zero since for every w there is a $-w$; the absolutely convergent w-series $\Sigma'\, 1/w^{2n}$ will be denoted by s_{2n}. It is a constant depending on ω_1 and ω_2; how it can be calculated will appear presently.

We have then

$$\zeta u = 1/u - s_4 u^3 - s_6 u^5 - \dots \qquad \dots\dots(5).$$

Hence $$\log \sigma u = \log u - s_4 u^4/4 - s_6 u^6/6 - \dots,$$

no constant being added since $\lim\limits_{u=0} \sigma u/u$ is 1;

and $$\sigma u = u \exp\left[-s_4 u^4/4 - s_6 u^6/6 - \dots\right]$$
$$= u\left[1 - s_4 u^4/4 - s_6 u^6/6 + \dots\right] \qquad \dots\dots(6).$$

We know from the factor-formula (§ 129) that this power series for σu has an infinite radius of convergence.

Further we have

$$\wp u = -\zeta' u = 1/u^2 + 3s_4 u^2 + 5s_6 u^4 + \dots \qquad \dots\dots(7),$$
$$\wp' u = -2/u^3 + 2\,.\,3s_4 u + 4\,.\,5s_6 u^3 + \dots \qquad \dots\dots(8).$$

About a point c which is not an infinity these functions can be developed in Taylor's series by the formula

$$fu = fc + (u-c)f'c + \frac{(u-c)^2}{2\,!}f''c + \dots,$$

this series holding good in a circle with centre c which extends up to the nearest infinity, in particular in the case of σu the series applies for every finite value of u.

131. Double Periodicity. The periodicity of $\wp' u$ is an evident consequence of the formula

$$\wp' u = -2\Sigma\, 1/(u-w)^3;$$

for if we write $u + w_0$ for u, where w_0 is any specified value of w, we only interchange the terms on the right and this has no effect on the series since it is absolutely convergent. Thus the function $\wp' u$ is periodic; it repeats, namely, its value when the argument u is increased by any one of the constants w. Because the constants w are built up from two independent numbers $2\omega_1$, $2\omega_2$ the function $\wp' u$ is called *doubly periodic*. More generally any one-valued analytic function fu which enjoys the property $f(u+w) = fu$ is doubly periodic*; if in addition it is a transcendental fractional function, that is, if it is one-valued and has no essential singular point in the finite part of the plane, it is called *elliptic*. Thus $\wp' u$ is an elliptic function.

It is clear that all the periods of $\wp' u$ are of the form $2m_1\omega_1 + 2m_2\omega_2$, where m_1, m_2 are *integers*. For since the equation

$$\wp'(u+a) = \wp' u$$

holds for all values of u, a is an infinity of $\wp' u$, as is seen by putting $u = 0$. But the only infinities of $\wp' u$ are the points w; hence a is one of the points w.

For this reason $2\omega_1$, $2\omega_2$ are called a primitive pair of periods of $\wp' u$:—not *the* primitive pair because (§ 127) they can be replaced by others which would give the same network of points w.

* We refer to other works for the proof that a one-valued analytic function cannot have more than two periods.

The property of $\wp' u$ that its periods form a network is one possessed by all elliptic functions fu. This means that elliptic functions have two and not more than two independent periods of which all other periods are multiples.

An elliptic function takes the same value at all congruent points. Thus if we know one point (say u_0) for which $\wp' u$ has a given value (say 0), we know further that $\wp' u$ is 0 at all the points $u_0 + w$. This of course does not imply conversely that all the points at which $\wp' u$ is 0 are included among the set $u_0 + w$; what it means is that the zeros arrange themselves in one or more networks. This peculiarity of elliptic functions,—that the points at which they take an assigned value fall into a finite number of networks,—is characteristic. It suffices for many purposes to consider such functions in a cell (§ 127) instead of over the whole plane. For the behaviour of the function in the cell whose corners are

$$u_0,\ u_0 + 2\omega_1,\ u_0 + 2\omega_1 + 2\omega_2,\ u_0 + 2\omega_2,$$

is repeated in every other congruent cell; and such cells cover the plane without overlapping. We shall denote such a cell by T.

It must be noticed that the cell is understood to be an open region. When we include one side we do not include the congruent opposite side; when we include one corner we exclude the other corners.

132. The Zeros of $\wp' u$. Since $\wp' u = \wp'(u + w)$, we have

$$\wp'(w - u) = \wp'(-u);$$

and therefore, since $\wp'$ is an odd function,

$$\wp'(w - u) = -\wp' u.$$

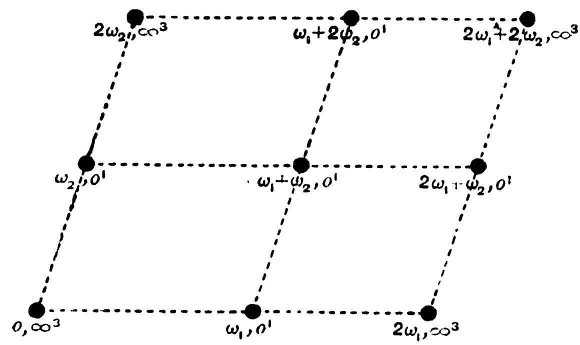
Let $u = w/2$; then

$$\wp'(w/2) = -\wp'(w/2),$$

whence $\qquad \wp'(w/2)$ is either 0 or ∞.

The points $w/2$ include all points $m_1\omega_1 + m_2\omega_2$; when m_1 and m_2 are both even this is simply the old network w, which we can call the network of periods; when m_1 and m_2 are not both even,

we have three new networks (fig. 59)* which can be called the networks of half-periods. At these latter points $\wp' u$ is zero, since it was ∞ at the network of periods and there only.



Zeros and infinities of $\wp' u$.

Fig. 59.

133. Are $\wp$, ζ, σ periodic? Having found that $\wp' u$ is periodic we naturally ask how the other functions $\wp$, ζ, σ stand in respect of periodicity. This could be answered from their defining formulæ (1), (2), (3); but it is more easily answered by means of integration.

Since $\wp' u$ is the derivate of $\wp u$, we have

$$\int_{u_0}^{u} \wp' u \, du = \wp u - \wp u_0,$$

the path of integration being any path which does not pass through a pole of $\wp' u$.

Observe how it is that the integral of $\wp' u$ is one-valued. At an infinity w of $\wp' u$ there is no residual term; that is, no term of the form $a/(u-w)$.

Hence also

$$\int_{u_0}^{u} \wp'(u+w)\, d(u+w) = \wp(u+w) - \wp(u_0+w),$$

that is, $$\int_{u_0}^{u} \wp'(u+w)\, du = \wp(u+w) - \wp(u_0+w).$$

* When we wish to represent, in a diagram, the values of a function at points of a plane, we write first the argument, or number attached to the point, second the corresponding value of the function.

But $\wp'(u+w)$ is always equal to $\wp' u$, therefore

$$\wp u - \wp u_0 = \wp(u+w) - \wp(u_0+w);$$

so that $\wp(u+w) - \wp u$ is a constant, k.

Putting $u = -w/2$ and observing that $\wp$ is an even function, we have

$$k = \wp(w/2) - \wp(-w/2) = 0.$$

Therefore $$\wp(u+w) = \wp u \quad\quad (9),$$

and $\wp u$ is also an elliptic function, with the same periods.

Let us apply the same argument to $\int \wp u\, du$. We have

$$\int_{u_0}^{u} \wp u\, du = -\zeta u + \zeta u_0,$$

$$\int_{u_0}^{u} \wp(u+w)\, du = -\zeta(u+w) + \zeta(u_0+w),$$

$$\zeta(u+w) - \zeta u = \text{a constant, } k.$$

Let $u = -w/2$. Then $k = \zeta(w/2) - \zeta(-w/2)$.

But ζ is odd; so that now $k = 2\zeta(w/2)$, and

$$\zeta(u+w) = \zeta u + 2\zeta(w/2).$$

Thus *ζu is not periodic*; it is said to be *quasi-periodic*, for by the addition of w the function is reproduced, save as to an added constant.

We have in particular

$$\zeta(u+2\omega_1) = \zeta u + 2\zeta\omega_1,$$

$$\zeta(u+2\omega_2) = \zeta u + 2\zeta\omega_2.$$

The constant $\zeta\omega_\lambda$, λ being always 1, 2, or 3, is denoted by η_λ. Thus

$$\zeta(u+2\omega_\lambda) = \zeta u + 2\eta_\lambda \quad\quad (10).$$

Hence

$$\begin{aligned}\zeta(u+2\omega_1+2\omega_2+2\omega_3) &= \zeta(u+2\omega_2+2\omega_3) + 2\eta_1 \\ &= \zeta(u+2\omega_3) + 2\eta_1 + 2\eta_2 \\ &= \zeta u + 2\eta_1 + 2\eta_2 + 2\eta_3.\end{aligned}$$

But by definition $2\omega_1 + 2\omega_2 + 2\omega_3$ is 0. Hence

$$\zeta u = \zeta u + 2\eta_1 + 2\eta_2 + 2\eta_3,$$

and therefore $$\eta_1 + \eta_2 + \eta_3 = 0 \quad\quad (11).$$

Again we have, applying the same procedure to $\int \zeta u du$,

$$\log \sigma u - \log \sigma u_0 = \int_{u_0}^{u} \zeta u \, du,$$

$$\log \sigma (u + w) - \log \sigma (u_0 + w) = \int_{u_0}^{u} \zeta (u + w) \, du,$$

$$\log [\sigma (u + w)/\sigma u] - \log [\sigma (u_0 + w)/\sigma u_0] = \int_{u_0}^{u} 2\zeta (w/2) \, du$$
$$= 2 (u - u_0) \zeta (w/2);$$

hence $\qquad \log [\sigma (u + w)/\sigma u] = 2u\zeta (w/2) + k,$

or $\qquad \sigma (u + w)/\sigma u = \exp [2u\zeta (w/2) + k].$

Let $u = - w/2$; then since σ is odd

$$- 1 = \exp [- w\zeta (w/2) + k],$$

and we find, by division, that

$$\sigma (u + w)/\sigma u = - \exp [(2u + w) \zeta (w/2)].$$

In particular

$$\sigma (u + 2\omega_\lambda) = - e^{2\eta_\lambda (u + \omega_\lambda)} \sigma u \dots\dots\dots\dots\dots (12).$$

Thus σu is not periodic; it is said to be *quasi-periodic*, because the function is reproduced save as to a simple exponential factor.

CHAPTER XIX.

ELLIPTIC FUNCTIONS.

134. Case of Liouville's Theorem. The existence of elliptic functions has been established by the actual construction of three functions of this kind, namely

$$\vartheta_3 u/\vartheta_2 u, \quad \wp u, \quad \wp' u\,;$$

we can at once extend the number. For clearly any rational algebraic function of $\wp u$, $\wp' u$ is a doubly periodic function and also a transcendental fractional function; or again the derivate of an elliptic function is another with the same periods.

The most important of all the *general* propositions employed in the discussion of elliptic functions is to this effect:—

No integral function, whether rational or transcendental, can be doubly periodic unless it reduce to a constant.

Let Gu be such a function; then $|Gu|$ is less than some fixed positive number γ when u is confined to a definite part of the u-plane, say within the parallelogram of periods (assumed to exist). The periodicity of Gu carries with it the inequality

$$|Gu| < \gamma$$

for all finite values of u. This result being at variance with the theorems of § 56 and § 93, the theorem is proved.

135. Integration round a Parallelogram. The structure of an elliptic function fu is discoverable from the behaviour of the function in a cell T (§ 131); Cauchy employed for this purpose the method of integration round the parallelogram

which bounds the cell. The spirit of this method will be apparent after a study of the following applications.

I. We suppose T so chosen that no side passes through a w. The integral $\int_T fudu$ taken round T is zero. For at congruent points on opposite sides of the parallelogram fu has the same value; so that

$$\int_{u_0}^{u_0+2\omega_1} fu\,du = \int_{u_0+2\omega_2}^{u_0+2\omega_1+2\omega_2} fu\,du,$$

whence
$$\int_{u_0}^{u_0+2\omega_1} fu\,du + \int_{u_0+2\omega_1+2\omega_2}^{u_0+2\omega_2} fu\,du = 0.$$

These are the integrals along the first and third sides. Similarly for the second and fourth sides

$$\int_{u_0+2\omega_1}^{u_0+2\omega_1+2\omega_2} fu\,du + \int_{u_0}^{u_0+2\omega_2} fu\,du = 0;$$

thus the whole result is zero. Hence, by § 119, the sum of the residues in T is zero; and it follows as an immediate consequence that *if fu has in T only one infinity, the infinity cannot be of the first order. But fu can have in T a single infinity, if the infinity is of the second or higher order; this was the case with $\wp u$.*

II. If gu be any polynomial, then from Cauchy's theory of residues (§ 120),

$$\int_T gu\,d\log fu = 2\pi i\,[\Sigma mgc - \Sigma m'gc'],$$

where c is a zero of fu of order m, c' an infinity of order m', and the summation is for all zeros and infinities in T. If gu be merely 1, then the formula is

$$\int_T f'u\,du/fu = 2\pi i\,(\Sigma m - \Sigma m').$$

But the same argument as before shows that the integral vanishes. Hence $\Sigma m = \Sigma m'$; that is, *in T the number of the zeros of fu is equal to the number of infinities, each zero or infinity being counted as often as its order indicates.*

Further the function $fu - k$ has in T the same number of zeros as infinities. But an infinity of $fu - k$ is an infinity of fu; hence

the number of points in T at which fu takes any assigned value k is independent of k. This number is *the order of the elliptic function*. Thus $\wp u$ is of the second order and has two zeros in a cell.

III. If instead of $gu = 1$ we take $gu = u$, we have

$$\int_T uf'udu/fu = 2\pi i\,(\Sigma mc - \Sigma m'c').$$

But $uf'u/fu$ takes at the points u and $u+2\omega_2$ values whose difference is $2\omega_2 f'u/fu$. Hence the integral along the first and third sides of T is

$$-\int_{u_0}^{u_0+2\omega_1} 2\omega_2 f'udu/fu$$

$$= -2\omega_2\,[\log f(u_0+2\omega_2) - \log fu_0]$$

$$= \text{a multiple of } 4\pi i\omega_2.$$

So for the second and fourth sides; whence

$$\Sigma mc - \Sigma m'c' = \text{some multiple of } 2\omega_1 \text{ and } 2\omega_2,$$

or

$$\Sigma mc \equiv \Sigma m'c';$$

or, in words, *the sum of the zeros is congruent with the sum of the infinities*. More generally, applying the same argument to $fu-k$, we can say that *the sum of the arguments for which an elliptic function takes an assigned value is congruent with the sum of its infinities.*

For the function $\wp u$ the theorem tells us nothing new; for we can infer from the even character of the function that the solutions of $\wp u = k$ pair off into opposite values u and $-u$.

IV. **Weierstrass's form of Legendre's Relation***. Let us integrate ζu round T. We have along the first and third sides

$$\int_{u_0}^{u_0+2\omega_1} \zeta u\,du - \int_{u_0}^{u_0+2\omega_1} \zeta(u+2\omega_2)\,du,$$

that is,

$$-\int_{u_0}^{u_0+2\omega_1} 2\eta_2\,du, \quad = -4\eta_2\omega_1.$$

* In Jacobi's notation this relation is $EK'+E'K-KK'=\pi/2$. See, for instance, A. C. Dixon's *Elliptic Functions*, p. 57.

Similarly along the other pair of sides we have $4\eta_1\omega_2$. Thus

$$\int_T \zeta u\,du = 4(\eta_1\omega_2 - \eta_2\omega_1).$$

On the other hand there is a simple infinity of ζu in T with a residue 1, but no other infinity. Thus the integral is $2\pi i$. Hence

$$\begin{vmatrix} \eta_1 & \eta_2 \\ \omega_1 & \omega_2 \end{vmatrix} = \pi i/2 \quad\ldots\ldots\ldots\ldots(13),$$

and similarly

$$\begin{vmatrix} \eta_2 & \eta_3 \\ \omega_2 & \omega_3 \end{vmatrix} = \begin{vmatrix} \eta_3 & \eta_1 \\ \omega_3 & \omega_1 \end{vmatrix} = \pi i/2.$$

In our choice of T we have made the chief amplitude of ω_2/ω_1 lie between 0 and π; that is, the coefficient of i in ω_2/ω_1 is positive. Had we taken the chief amplitude between 0 and $-\pi$ this coefficient would have been negative. In the one case the description of the parallelogram is positive, in the other negative; and therefore this second case would have given $-\pi i/2$, not $\pi i/2$.

We shall give an instance of the way in which this relation can be used to simplify formulæ in elliptic functions.

By successive applications of the formulæ

$$\sigma(u + 2\omega_1) = -e^{2\eta_1(u+\omega_1)}\sigma u,$$

$$\sigma(u + 2\omega_2) = -e^{2\eta_2(u+\omega_2)}\sigma u,$$

we get

$$\sigma(u + w)$$
$$= (-)^{m_1+m_2}\exp\left[2(m_1\eta_1 + m_2\eta_2)(u + w/2) + 2m_1m_2(\eta_1\omega_2 - \eta_2\omega_1)\right]\sigma u.$$

Using the formula for $\eta_1\omega_2 - \eta_2\omega_1$ this becomes

$$\sigma(u + w) = (-)^{m_1+m_2+m_1m_2}\exp 2\eta(u + w/2)\,.\,\sigma u \quad\ldots\ldots(14),$$

where $\qquad \eta = m_1\eta_1 + m_2\eta_2.$

136. Comparison of Elliptic Functions. Consider two elliptic functions f_1u and f_2u with the same primitive periods.

I. If they have the same zeros and infinities, in each case to the same order,—that is, if f_1u is ∞^m at c, so is f_2u and so on,—then f_1u/f_2u has no infinities. Each infinity of f_1u is cancelled by an infinity of f_2u and each zero of f_2u by a zero of f_1u. Hence (§ 134) the ratio is a constant.

Thus an elliptic function is determined save as to a constant factor when its zeros and infinities are assigned.

II. Again if the two elliptic functions $f_1 u$ and $f_2 u$ have the same infinities and the negative powers of the Laurent series about each common infinity are the same for both, their difference is an integral elliptic function; that is, by Liouville's theorem, a constant.

III. Let $f_1 u$ and $f_2 u$ be of the second order, with common infinities c and c' but different residues r_1 and r_2 at c. Then near c

$$f_1 u = r_1/(u-c) + P(u-c),$$
$$f_2 u = r_2/(u-c) + Q(u-c),$$

and therefore $r_2 f_1 u - r_1 f_2 u$ is finite at c. That is, it is an elliptic function with not more than one infinity in T, and that simple; it is therefore a constant.

137. Algebraic Equation connecting the Functions $\wp u$, $\wp' u$. Taking the second method of comparison of § 136,—that by infinities alone,—we know that near $u = 0$

$$\wp u = 1/u^2 + 3s_4 u^2 + 5s_6 u^4 + \ldots,$$
$$\wp' u = -2/u^3 + 2 \,.\, 3s_4 u + 4 \,.\, 5s_6 u^3 + \ldots.$$

Thus $\wp'^2 u$ is of the sixth order; and we have

$$\wp'^2 u = 4/u^6 - 24s_4/u^2 - 80s_6 + P_1 u^2,$$
$$\wp^3 u = 1/u^6 + 9s_4/u^2 + 15s_6 + P_1 u^2.$$

Hence $$\wp'^2 u - 4\wp^3 u = -60s_4/u^2 - 140s_6 + P_1 u^2,$$

and $$\wp'^2 u - 4\wp^3 u + 60s_4 \wp u = -140s_6 + P_1 u^2.$$

The function on the left is an elliptic function which has no longer an infinity at $u=0$; for on the right there are no negative powers of u. Hence it is a constant; and the constant, by letting u be zero, is seen to be $-140s_6$. Hence

$$\wp'^2 u = 4\wp^3 u - 60s_4 \wp u - 140s_6.$$

If $$\left.\begin{aligned} g_2 &= 60s_4 = 60\Sigma' 1/w^4, \\ g_3 &= 140s_6 = 140\Sigma' 1/w^6 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(15),$$

the equation is $$\wp'^2 u = 4\wp^3 u - g_2 \wp u - g_3 \ldots\ldots\ldots\ldots\ldots\ldots(16).$$

The numbers g_2 and g_3 are called the *invariants*.

They are invariants of the network because they are not altered when we replace $2\omega_1$ and $2\omega_2$,—the original given constants,—by another primitive pair; they are also invariants of the expression $4\wp^3 - g_2\wp - g_3$ regarded as a quartic in $\wp$ one of whose zeros is $\wp = \infty$.

138. The Addition Theorem for $\wp u$. Let us apply the theorem that the sum of the zeros of an elliptic function is congruent with the sum of the infinities, to express $\wp(u_1 + u_2)$ in terms of functions of u_1 and u_2 alone.

Consider $\wp' u - c\wp u - c'$, where c and c' are constants. This is an elliptic function with the same periods as $\wp u$; it is ∞ in T only at $u = 0$, and is ∞ there like $-2/u^3 - c/u^2$. It is therefore of the third order and has three zeros u_1, u_2, u_3 in T. The theorem states then that

$$u_1 + u_2 + u_3 \equiv 0.$$

Now since for each of these points

$$(c\wp u + c')^2 = (\wp' u)^2 = 4\wp^3 u - g_2\wp u - g_3,$$

we have
$$\wp u_1 + \wp u_2 + \wp u_3 = c^2/4$$
$$= \frac{1}{4}\left(\frac{\wp' u_1 - \wp' u_2}{\wp u_1 - \wp u_2}\right)^2,$$

and since $\wp u$ is an even function,

$$\wp(u_1 + u_2) = \wp u_3;$$

therefore
$$\wp(u_1 + u_2) = \frac{1}{4}\left(\frac{\wp' u_1 - \wp' u_2}{\wp u_1 - \wp u_2}\right)^2 - \wp u_1 - \wp u_2 \quad \text{.........(17)},$$

which is the *addition theorem* of the $\wp$-function. By differentiation with respect to u_1 we obtain an addition theorem for $\wp'(u_1 + u_2)$.

We know that $u = \omega_1, \omega_2, \omega_3$ are zeros of $\wp' u$.

Let $\wp\omega_\lambda = e_\lambda$. Then $4\wp^3 u - g_2\wp u - g_3$ vanishes when $\wp u = e_\lambda$; thus it is, in factors,

$$4(\wp u - e_1)(\wp u - e_2)(\wp u - e_3),$$

whence
$$\left.\begin{aligned} e_1 + e_2 + e_3 &= 0, \\ e_2e_3 + e_3e_1 + e_1e_2 &= -g_2/4, \\ e_1e_2e_3 &= g_3/4 \end{aligned}\right\} \quad \text{.................(18)}.$$

And we have of course

$$\wp'^2 u = 4(\wp u - e_1)(\wp u - e_2)(\wp u - e_3) \quad \text{.........(10)}.$$

From (16) we derive

$$2\wp' u \wp'' u = 12\wp^2 u \wp' u - g_2 \wp' u,$$

or

$$\wp'' u = 6\wp^2 u - g_2/2,$$

$$\wp''' u = 12\wp u \wp' u,$$

and so on; the $2n$th derivate is a polynomial in $\wp u$ of degree $n+1$, and the $(2n+1)$th derivate is the product of $\wp' u$ into a polynomial in $\wp u$ of degree n.

If in any of these,—the last written is the most convenient,—we substitute the series (7) for $\wp u$ and the derived series for $\wp' u$, $\wp'' u$, we infer, by equating coefficients, a recurrence formula for s_{2n} in terms of $s_4, s_6, \ldots, s_{2n-2}$. Hence it follows that s_{2n} is a rational function of g_2 and g_3; that is, all the numbers $\Sigma' 1/w^{2n}$, where $n > 1$, can be expressed rationally by means of the first two.

Ex. Prove $\Sigma' 1/w^8 = g_2^2/2^4 \cdot 3 \cdot 5^2 \cdot 7$, $\Sigma' 1/w^{10} = g_2 g_3/2^4 \cdot 3 \cdot 5 \cdot 7 \cdot 11$.

Some special cases of the addition theorem must be noted. First let $u_2 = \omega_\lambda$; then $\wp' \omega_\lambda = 0$, $\wp \omega_\lambda = e_\lambda$; and the formula becomes, on replacing u_1 by u,

$$\begin{aligned}
\wp(u + \omega_\lambda) &= \frac{1}{4}\left(\frac{\wp' u}{\wp u - e_\lambda}\right)^2 - \wp u - e_\lambda \\
&= \frac{(\wp u - e_\mu)(\wp u - e_\nu)}{\wp u - e_\lambda} - \wp u - e_\lambda \\
&= \frac{e_\lambda \wp u + e_\mu e_\nu + e_\lambda^2}{\wp u - e_\lambda} \\
&= e_\lambda + \frac{e_\mu e_\nu + 2e_\lambda^2}{\wp u - e_\lambda};
\end{aligned}$$

or

$$\begin{aligned}
[\wp(u + \omega_\lambda) - e_\lambda][\wp u - e_\lambda] &= e_\mu e_\nu + 2e_\lambda^2 \\
&= (e_\mu - e_\lambda)(e_\nu - e_\lambda) \quad \ldots\ldots(20).
\end{aligned}$$

Second let $u_2 = u_1 = u$. Then

$$\begin{aligned}
\wp 2u + 2\wp u &= \lim \frac{1}{4}\left(\frac{\wp' u_1 - \wp' u_2}{\wp u_1 - \wp u_2}\right)^2 \\
&= \frac{1}{4}\left(\frac{\wp'' u}{\wp' u}\right)^2 \\
&= \frac{(3\wp^2 - g_2/4)^2}{4\wp^3 - g_2\wp - g_3};
\end{aligned}$$

or
$$\wp 2u = \frac{\wp^4 + \frac{1}{2}g_2\wp^2 + 2g_3\wp + \frac{1}{16}g_2^2}{4\wp^3 - g_2\wp - g_3}.$$

If again we put in the addition theorem $u_1 = 2u$, $u_2 = u$, we can obtain a formula which when reduced gives $\wp 3u$. But the proper method for obtaining $\wp nu$ in terms of $\wp u$ is by means of the function $\sigma nu/\sigma^{n^2}u$.

In the addition theorem there occurs the square of the expression
$$\frac{1}{2}\frac{\wp' u_1 - \wp' u_2}{\wp u_1 - \wp u_2}.$$

This expression is to be remarked, as it is often met with in the study of elliptic functions. To view it as a function of one variable only let us write it
$$\frac{1}{2}\frac{\wp' u - \wp' a}{\wp u - \wp a}.$$

It is an elliptic function with the same periods as $\wp u$; it is ∞^1 at $u = 0$, with the residue -1; it is not ∞ at $u = a$, since both numerator and denominator have a simple zero there. But it is ∞^1 at $u = -a$; and the residue there is $+1$. It is then an elliptic function of order 2 with one infinity at $u = -a$ but the other at $u = 0$ and independent of a.

Changing the sign of a the function
$$\frac{1}{2}\frac{\wp' u + \wp' a}{\wp u - \wp a}$$
has infinities at a, 0 with residues 1, -1.

139. Expression of an Elliptic Function by means of $\wp u$. Taking now any elliptic function fu with the same periods as $\wp u$ let it be ∞ at an infinity a like
$$c_{-1}/(u-a) + c_{-2}/(u-a)^2 + \ldots + c_{-m}/(u-a)^m.$$
The functions $\wp(u-a)$, $\wp'(u-a)$,... are ∞ at a like
$$1/(u-a)^2,\ -2/(u-a)^3, \ldots$$
Hence by subtracting
$$c_{-2}\wp(u-a),\ -\frac{1}{2}c_{-3}\wp'(u-a),\ \frac{1}{3!}c_{-4}\wp''(u-a), \text{ etc.},$$

we remove all the infinite terms except the first. To remove this one, we subtract

$$\frac{1}{2} c_{-1} \frac{\wp' u + \wp' a}{\wp u - \wp a}.$$

But in so doing we introduce a simple infinity at the origin. When however we remove in this way all the infinite terms at all the infinities of fu, there remains an elliptic function with at most one network of simple infinities; that is, a constant k.

Thus we have, for any elliptic function, an expression

$$\begin{aligned} fu = k + \Sigma \frac{1}{2} c_{-1} & \frac{\wp' u + \wp' a}{\wp u - \wp a} \\ & + \Sigma c_{-2} \wp (u - a) \\ & - \tfrac{1}{2} \Sigma c_{-3} \wp' (u - a) - \ldots \ldots\ldots\ldots\ldots \text{(i)}. \end{aligned}$$

From the addition theorem, $\wp (u - a)$ and its derivates are rational functions of $\wp u$ and $\wp' u$; hence we can express fu itself as a rational function of $\wp u$ and $\wp' u$. If fu is even the odd function $\wp' u$ will disappear, but if fu is odd $\wp' u$ will occur as a factor, all higher powers of $\wp' u$ being removed by the fundamental formula (16). And generally we can say that

$$fu = R_1 (\wp u) + \wp' u R_2 (\wp u),$$

where R_1 and R_2 mean 'rational function of.'

Hence *any two elliptic functions $f_1 u$ and $f_2 u$ with the same periods are connected by an algebraic equation.* For we have the three equations

$$\begin{aligned} f_1 u &= R_1 (\wp u) + \wp' u R_2 (\wp u), \\ f_2 u &= R_3 (\wp u) + \wp' u R_4 (\wp u), \\ \wp'^2 u &= 4\wp^3 - g_2 \wp - g_3, \end{aligned}$$

from which we can eliminate $\wp u$ and $\wp' u$. In particular an elliptic function fu and its derivate $f'u$, having the same periods, are connected by an algebraic relation; an instance of this is the formula (16).

140. The Addition Theorem for ζu. We have said that ζ is not elliptic. But such an expression as $\zeta(u + a) - \zeta u$ is elliptic; for

$$\begin{aligned} & \zeta (u + 2\omega_\lambda + a) - \zeta (u + 2\omega_\lambda) \\ &= \zeta (u + a) + 2\eta_\lambda - \zeta u - 2\eta_\lambda \\ &= \zeta (u + a) - \zeta u. \end{aligned}$$

The infinities are $u \equiv 0$, $u \equiv -a$; and the residues are respectively -1 and $+1$.

The function
$$\frac{1}{2}\frac{\wp' u - \wp' a}{\wp u - \wp a}$$
has the same infinities and residues. Hence
$$\zeta(u+a) - \zeta u - k = \frac{1}{2}\frac{\wp' u - \wp' a}{\wp u - \wp a}.$$
To determine the constant we expand near the origin. The expression on the left is
$$\zeta a + u\zeta' a - 1/u + P_2 u - k,$$
and that on the right is
$$\frac{1}{2u}\frac{-2 - u^3\wp' a + P_2 u^2}{1 - u^2\wp a + P_2 u^2},$$
$$= -\frac{1}{u}\left[1 + \frac{u^3}{2}\wp' a + \ldots\right][1 + u^2\wp a + \ldots],$$
$$= -1/u - u\wp a + \ldots.$$
Comparing the constant terms we have $k = \zeta a$. Hence the *addition theorem*:
$$\zeta(u+a) - \zeta u - \zeta a = \frac{1}{2}\frac{\wp' u - \wp' a}{\wp u - \wp a} \quad \ldots\ldots\ldots\ldots\ldots(21).$$

141. Integration of an Elliptic Function. Integrating the formula (1) of § 139 we have
$$\int f u du = k' + ku + \Sigma \frac{1}{2} c_{-1} \int \frac{\wp' u + \wp' a}{\wp u - \wp a} du$$
$$- \Sigma c_{-2}\zeta(u-a) - \tfrac{1}{2}\Sigma c_{-3}\wp(u-a) + \ldots$$
$$+ (-)^m \frac{1}{(m-1)!}\Sigma c_{-m}\wp^{(m-3)}(u-a).$$
We have then, for the integral of any elliptic function, (1) a term ku, (2) terms of the form $\zeta(u-a)$, (3) integrals of the form
$$\frac{1}{2}\int \frac{\wp' u + \wp' a}{\wp u - \wp a} du,$$
(4) elliptic functions. By the addition theorem for ζu,
$$\zeta(u-a) = \zeta u - \zeta a + \frac{1}{2}\frac{\wp' u + \wp' a}{\wp u - \wp a};$$

so that the terms (2) contribute a single term in ζu. Also integrating the last equation we have

$$\frac{1}{2}\int \frac{\wp' u + \wp' a}{\wp u - \wp a}\, du = \log \frac{\sigma(u-a)}{\sigma u} + u\zeta a.$$

Thus on the whole the integral is

$$\int fu\, du = k_1 u - \Sigma c_{-2}\zeta u + \Sigma c_{-1} \log \frac{\sigma(u-a)}{\sigma u} + f_1 u,$$

where $f_1 u$ is an elliptic function.

Ex. Prove that $\int_{\omega_2}^{\omega_2+2\omega_1} \wp'^2 u\, du = \frac{2}{5}(2g_2\eta_1 - 3g_3\omega_1)$, u being arbitrary.

142. Expression of an Elliptic Function by means of σu. We shall next show how the theory of elliptic functions can be deduced directly from the σ-function without the intervention of the $\wp$-function, and without the use of integration. We shall thus have alternative proofs of some of the preceding theorems.

I. *An elliptic function fu has a finite number of zeros and a finite number of infinities in a cell.*

The reasoning of § 104 excludes the possibility of there being infinitely many non-essential singularities; for such a system of points has limit-points and these limit-points are essential singularities, whereas fu has no essential singularity except at $u = \infty$. And similarly by § 106 fu cannot have infinitely many zeros within the parallelogram.

II. *Every elliptic function fu can be expressed in the form*

$$e^{Gu}\frac{\sigma(u-a_1)\,\sigma(u-a_2)\,\ldots\,\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\,\ldots\,\sigma(u-b_s)},$$

where the a's are the zeros and the b's are the infinities of fu in the cell.

Each factor in the numerator of

$$\frac{\sigma(u-a_1)\,\sigma(u-a_2)\,\sigma(u-a_3)\,\ldots\,\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\,\sigma(u-b_3)\,\ldots\,\sigma(u-b_s)}$$

vanishes at the corresponding a and at all points congruent to a; hence the zeros of the numerator are the same as the zeros of fu. Similarly the infinities of the quotient, or the zeros of the denominator, are the infinities of fu. Hence the quotient of fu

by the expression in the σ's is a one-valued analytic function which has neither zeros nor infinities; this quotient must therefore be of the form e^{Gu} (§ 93) and the theorem is proved.

It is not true conversely that every expression of the form

$$e^{Gu}\frac{\sigma(u-a_1)\,\sigma(u-a_2)\,\sigma(u-a_3)\dots\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\,\sigma(u-b_3)\dots\sigma(u-b_s)}$$

represents an elliptic function. We shall prove, for example, that the number of the a's must be equal to the number of the b's, and that Gu must be a special polynomial which reduces, under certain conditions, to a constant.

III. *The number of zeros of an elliptic function fu in the cell is equal to the number of infinities in the same region.*

We suppose that the zeros and infinities are all simple. This does not interfere with the generality of the results, for the case in which a zero (or infinity) is multiple is merely a limiting case in which several points previously distinct have moved into coincidence.

Let $a_1, a_2, \dots, a_r$ and $b_1, b_2, \dots, b_s$ be, as before, the zeros and infinities that are situated within the parallelogram of periods. Then

$$fu = e^{Gu}\frac{\sigma(u-a_1)\,\sigma(u-a_2)\dots\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\dots\sigma(u-b_s)}.$$

The expression on the right-hand side is to be reproduced when u is changed into $u+2\omega_\lambda$; hence since

$$\sigma(u-a+2\omega_\lambda) = -e^{2\eta_\lambda(u-a+\omega_\lambda)}\sigma(u-a),$$

we must have

$$1 = (-)^{r-s}\exp\left[G(u+2\omega_\lambda)-Gu\right].$$
$$\exp\left[2(r-s)\eta_\lambda(u+\omega_\lambda)-2\eta_\lambda\left(\sum_{n=1}^{r}a_n-\sum_{n=1}^{s}b_n\right)\right]$$

This requires that the expression

$$2(r-s)\,\eta_\lambda(u+\omega_\lambda)-2\eta_\lambda\left(\sum_{n=1}^{r}a_n-\sum_{n=1}^{s}b_n\right)+G(u+2\omega_\lambda)-Gu,$$

shall be a multiple of πi. Hence

$$G(u+2\omega_\lambda)-Gu$$
$$=2\eta_\lambda\left(\sum_{n=1}^{r}a_n-\sum_{n=1}^{s}b_n\right)-2(r-s)\,\eta_\lambda(u+\omega_\lambda)+h_\lambda\pi i\dots\dots\text{(i)}.$$

This integer h_λ must be the same for all values of u; for the expressions in u change continuously, whereas h_λ if it change at all can only change discontinuously.

By differentiating (i) twice we get

$$G''(u+2\omega_\lambda) = G''u;$$

a relation which shows that the transcendental integral function $G''u$ is doubly periodic. It can have no infinity within the parallelogram of periods and must reduce therefore to a constant. Hence

$$Gu = A_0 u^2 + A_1 u + A_2.$$

Substituting this value in (i) and putting $\lambda = 1, 2$ successively,

$$4A_0\omega_1(u+\omega_1) + 2A_1\omega_1$$
$$= 2\eta_1\left(\sum_{n=1}^{r} a_n - \sum_{n=1}^{s} b_n\right) - 2\eta_1(r-s)(u+\omega_1) + h_1\pi i \ldots\text{(ii)},$$

$$4A_0\omega_2(u+\omega_2) + 2A_1\omega_2$$
$$= 2\eta_2\left(\sum_{n=1}^{r} a_n - \sum_{n=1}^{s} b_n\right) - 2\eta_2(r-s)(u+\omega_2) + h_2\pi i \ldots\text{(iii)}.$$

As these equations are to be satisfied by all values of u, we can equate the terms in u on the two sides of (ii) and (iii); hence

$$4A_0\omega_1 = -2\eta_1(r-s),\quad 4A_0\omega_2 = -2\eta_2(r-s) \ldots\ldots\text{(iv)}.$$

Suppose that in the equations (iv), $r - s \neq 0$; then

$$\eta_1\omega_2 - \eta_2\omega_1 = 0,$$

a result at variance with the formula (13) of § 135. Hence r must be equal to s, a theorem established already by means of integration round the parallelogram.

The preceding equations give a new proof of the following theorem :—

IV. *The sum of the zeros of an elliptic function in a cell is equal to the sum of the infinities in the same region increased by a suitable period.*

By equating those terms of (ii) and (iii) that are independent of u, we get

$$A_1\omega_1 = \eta_1\left(\sum_{n=1}^{r} a_n - \sum_{n=1}^{r} b_n\right) + h_1\pi i,$$

$$A_1\omega_2 = \eta_2\left(\sum_{n=1}^{r} a_n - \sum_{n=1}^{r} b_n\right) + h_2\pi i.$$

Hence $$(\eta_1\omega_2 - \eta_2\omega_1)\left(\sum_{n=1}^{r} a_n - \sum_{n=1}^{r} b_n\right) = (h_2\omega_1 - h_1\omega_2)\,\pi i,$$

$$(\eta_1\omega_2 - \eta_2\omega_1)\,A_1 = (h_2\eta_1 - h_1\eta_2)\,\pi i.$$

But $\eta_1\omega_2 - \eta_2\omega_1 = \pi i/2$; hence

$$\sum_{n=1}^{r} a_n - \sum_{n=1}^{r} b_n = 2\,(h_2\omega_1 - h_1\omega_2),$$

and $$A_1 = 2\,(h_2\eta_1 - h_1\eta_2).$$

Hence the theorem is proved; and further it is proved that *the most general form for an elliptic function is*

$$A\,\frac{\sigma(u-a_1)\,\sigma(u-a_2)\ldots\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\ldots\sigma(u-b_r)}\,e^{(2h_2\eta_1-2h_1\eta_2)u} \quad \ldots(22),$$

where h_1, h_2 are integers given by

$$\sum_{n=1}^{r} a_n - \sum_{n=1}^{r} b_n = 2\,(h_2\omega_1 - h_1\omega_2).$$

This general formula shows that two elliptic functions with the same zeros and the same infinities in the parallelogram of periods differ merely by a constant factor (§ 136).

V. *The equation $fu = k$ admits r non-congruent roots, and the sum of these roots is equal to the sum of the infinities of fu increased by a suitable period.*

This is proved at once by applying Theorems III., IV. to the elliptic function $fu - k$.

143. Relation connecting $\wp u$, σu. The formula

$$fu = A e^{2(h_2\eta_1 - h_1\eta_2)u}\,\frac{\sigma(u-a_1)\,\sigma(u-a_2)\ldots\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\ldots\sigma(u-b_r)}$$

is simplified if we replace a_1 by the congruent number

$$a'_1 = a_1 - 2h_2\omega_1 + 2h_1\omega_2.$$

It then runs

$$fu = A'\,\frac{\sigma(u-a'_1)\,\sigma(u-a_2)\ldots\sigma(u-a_r)}{\sigma(u-b_1)\,\sigma(u-b_2)\ldots\sigma(u-b_r)},$$

and now the sum of the a's is equal to the sum of the b's.

Let us apply this formula to the function

$$\wp u - \wp v,$$

where v is treated as a constant. The number r is now 2, the two infinities b are 0, 0, the two zeros a are $v, -v$. Hence

$$\wp u - \wp v = A' \frac{\sigma(u+v)\,\sigma(u-v)}{\sigma^2 u}$$

To determine A' multiply both sides by u^2 and let u tend to zero. We get $\lim\limits_{u=0} u^2 \wp u = 1$ and $\lim\limits_{u=0} \{u/\sigma u\}^2 = 1$. Hence

$$1 = A'(-\sigma^2 v),\ A' = -1/\sigma^2 v.$$

Thus $$\wp u - \wp v = -\frac{\sigma(u+v)\,\sigma(u-v)}{\sigma^2 u\,\sigma^2 v} \quad \text{............(23)}.$$

By logarithmic differentiation of this formula with respect to u and v, we get

$$\zeta(u+v) + \zeta(u-v) - 2\zeta u = \frac{\wp' u}{\wp u - \wp v},$$

$$\zeta(u+v) - \zeta(u-v) - 2\zeta v = \frac{-\wp' v}{\wp u - \wp v}.$$

By adding and subtracting these two formulæ we get

$$\zeta(u \pm v) = \zeta u \pm \zeta v + (\wp' u \mp \wp' v)/2(\wp u - \wp v).$$

This is the addition theorem for ζu.

Ex. Deduce by differentiation the addition theorem for $\wp(u+v)$.

144. The Functions $\sqrt{\wp u - e_\lambda}$. Writing in (23) $v = \omega_\lambda$ we get

$$\wp u - e_\lambda = \sigma(\omega_\lambda + u)\,\sigma(\omega_\lambda - u)/\sigma^2\omega_\lambda \sigma^2 u,$$

and from (12) we have

$$\sigma(\omega_\lambda + u) = e^{2\eta_\lambda u}\sigma(\omega_\lambda - u).$$

Hence $$\wp u - e_\lambda = [e^{-\eta_\lambda u}\sigma(\omega_\lambda + u)/\sigma\omega_\lambda \sigma u]^2.$$

Thus the square roots of $\wp u - e_\lambda$ are distinct one-valued functions.

Ex. Explain this by considering the zeros and infinities of $\sqrt{\wp u - e_\lambda}$, as in § 99, Ex. 3.

We write $\sqrt{\wp u - e_\lambda} = e^{-\eta_\lambda u}\sigma(\omega_\lambda + u)/\sigma\omega_\lambda \sigma u$ ....(24), the square root selected being that for which

$$\lim u\sqrt{\wp u - e_\lambda} = 1 \text{ when } u = 0.$$

This function $\sqrt{\wp u - e_\lambda}$ is odd; it has a simple zero when

$u \equiv \omega_\lambda$, a simple infinity when $u \equiv 0$; it can at most change sign when u increases by a period, so that it is elliptic. To determine its periods let u become $u + 2\omega_\mu$; then $\sigma(\omega_\lambda + u)/\sigma u$ is multiplied by $\exp 2\eta_\mu\omega_\lambda$, and therefore $\sqrt{\wp u - e_\lambda}$ by $\exp 2(\eta_\mu\omega_\lambda - \eta_\lambda\omega_\mu)$, that is, by 1 when μ is λ, and by $\exp \pm \pi i$, or -1 when μ is not λ. Hence any two of $2\omega_\lambda$, $4\omega_\mu$, $4\omega_\nu$, are primitive periods.

Ex. Writing $\sqrt{\wp u - e_\lambda} = \xi_\lambda$, prove the differential equation

$$\xi'^2_\lambda = (\xi_\lambda^2 + e_\lambda - e_\mu)(\xi_\lambda^2 + e_\lambda - e_\nu).$$

145. Connexion of the Functions $\wp u$ and $\vartheta_3 v/\vartheta_2 v$. We are now in a position to connect the elliptic function $\wp u$ whose periods are $2\omega_1$, $2\omega_2$ with the elliptic function $\vartheta_3 v/\vartheta v$ whose periods are 1 and 2ω.

Let $u = 2\omega_1 v$; then $\wp u$ becomes $\wp 2\omega_1 v$ with periods $v = 1$ and $v = \omega_2/\omega_1$. Let $\omega_2/\omega_1 = \omega$. The function $\vartheta_3 v/\vartheta v$ has zeros congruent with $-(1+\omega)/2$, and infinities congruent with $\omega/2$. The function $\sqrt{\wp 2\omega_1 v - e_3}$ has zeros congruent with $\omega_3/2\omega_1$, that is, with $-(1+\omega)/2$, and infinities congruent with 0.

The function $\sqrt{\wp 2\omega_1 v - e_2}$ has the same infinities, and its zeros are congruent with $\omega/2$. Hence $\sqrt{\wp 2\omega_1 v - e_3}/\sqrt{\wp 2\omega_1 v - e_2}$ has the same zeros and infinities as $\vartheta_3 v/\vartheta_2 v$. Hence the ratio of the two is a constant which, letting v tend to zero, is ϑ_2/ϑ_3, where ϑ_λ is the value of $\vartheta_\lambda v$ when v is 0. Hence finally

$$\sqrt{\wp 2\omega_1 v - e_3}/\sqrt{\wp 2\omega_1 v - e_2} = \vartheta_3 v \vartheta_2/\vartheta_2 v \vartheta_3 \;\ldots\ldots\; (25).$$

Ex. 1. By writing $v = 1/2$ prove that

$$(e_1 - e_3)/(e_1 - e_2) = \vartheta_2^4/\vartheta_3^4.$$

Ex. 2. By expanding in powers of v prove that

$$4\omega_1^2(e_2 - e_3) = \vartheta_3''/\vartheta_3 - \vartheta_2''/\vartheta_2,$$

where ϑ_λ'' is the value of $D_v^2\vartheta_\lambda v$ when $v = 0$.

146. Connexion of the Functions σu and ϑv. When u increases by $2\omega_1$, σu is multiplied by $-\exp 2\eta_1(u + \omega_1)$. The function $\exp(\eta_1 u^2/2\omega_1)$, in the same case, is multiplied by $\exp 2\eta_1(u + \omega_1)$. Hence the function $\sigma u \exp(-\eta_1 u^2/2\omega_1)$ merely changes sign when u becomes $u + 2\omega_1$; therefore it has the period $4\omega_1$. When u becomes $u + 2\omega_2$, the ratio is multiplied

by $$-\exp\left[2\eta_2(u+\omega_2)-2\eta_1\omega_2(u+\omega_2)/\omega_1\right],$$
or since $$\eta_1\omega_2-\eta_2\omega_1=\pi i/2,$$
by $$-\exp\left[-\pi i(u+\omega_2)/\omega_1\right].$$
Write for shortness
$$u=2\omega_1 v,\ z=e^{\pi i v};\ \omega_2/\omega_1=\omega,\ q=e^{\pi i\omega};$$
then the function $\sigma(2\omega_1 v)\exp-2\eta_1\omega_1 v^2$ has the period 2.

Therefore the function of v, when placed on the z-plane, is *one-valued*; its singular points arise only from $v=\infty$, and are $z=0$, $z=\infty$. Thus (§ 123) there is for the function a Laurent series $\sum\limits_{-\infty}^{\infty} a_n z^n$. Since the function of v is odd, the Laurent series changes sign when we write $1/z$ for z, and is $\sum\limits_{1}^{\infty} c_n(z^n-z^{-n})$.

Calling this series ϕz we have to determine the coefficients. Now when u becomes $u+2\omega_2$, v becomes $v+\omega$, z becomes qz, and the function is multiplied by $-1/qz^2$. Thus
$$\phi(qz)=-\phi z/qz^2.$$
This functional equation will enable us to determine the coefficients; for we have already found in § 125 the Laurent series which satisfies the functional equation $xf(q^2x)=-fx$, or $z^2f(q^2z^2)=-fz^2$, an equation which can be easily transformed into the functional equation for ϕz. For let $fz^2=kz\phi z$; then $f(q^2z^2)=kqz\phi(qz)$, whence $qz^2\phi(qz)=-\phi z$.

Thus the function is fz^2/kz where k is an arbitrary constant, and f has the same meaning as in § 125; hence also (§ 126) $\phi z=k_1\vartheta v$, or
$$\sigma 2\omega_1 v=k_1\vartheta v\exp 2\eta_1\omega_1 v^2.$$
Expanding in powers of v, we have
$$\sigma 2\omega_1 v=2\omega_1 v+P_3 v,\ \vartheta v=v\vartheta'0+v^3\vartheta'''0/6+\ldots,$$
$$\exp 2\eta_1\omega_1 v^2=1+2\eta_1\omega_1 v^2+\ldots;$$
so that equating coefficients
$$2\omega_1=k_1\vartheta'0,\ 2\eta_1\omega_1\vartheta'0+\vartheta'''0/6=0.$$
Hence finally, eliminating k_1, we get
$$\sigma 2\omega_1 v=2\omega_1\exp 2\eta_1\omega_1 v^2.\vartheta v/\vartheta'0 \qquad (26),$$
and $$12\eta_1\omega_1=-\vartheta'''0/\vartheta'0 \qquad (27).$$

These equations with those of the preceding article are of great use in numerical applications of elliptic functions; and it is for this end that we have selected them as illustrations. When the network is given, that primitive pair of periods is to be selected which makes $q, = \exp i\omega_2/\omega_1$, as small as possible; the q-series then converge with remarkable rapidity. The formulæ given as examples in § 145, together with $e_1 + e_2 + e_3 = 0$, suffice for the calculation of e_1, e_2, e_3 in terms of ω_1; formula (27) gives η_1, (13) gives η_2, and (26) gives σu. The value of $\wp u$ can be calculated either by means of its expression $-D^2 \log \sigma u$ or from (25).

On the other hand, (26) gives the expression of ϑv in primary factors (ch. XV.), from which in turn we can infer the analogous expressions for the other ϑ-functions $\vartheta_\lambda v$.

CHAPTER XX.

SIMPLE ALGEBRAIC FUNCTIONS ON RIEMANN SURFACES.

147. **The Square Root.** In the present chapter we study some simple algebraic functions with the help of the apparatus known as the Riemann surface.

Example I. $y^2 = x$. We begin with the square root of x. Let $y^2 = x$, and let x, y be $\rho e^{i\theta}$, $\rho' e^{i\theta'}$ respectively. Then

$$\rho' = \rho^{1/2} \text{ and } \theta' \equiv \theta/2, \text{ or } \theta/2 + \pi \text{ (mod. } 2\pi).$$

There are accordingly two *opposite* values of y for an assigned x. Taking one of these arbitrarily at an initial point and then making x describe a continuous path which passes through neither 0 nor ∞, we can assign one and only one value y to each point of the path in such a way that these y's shall vary continuously as x describes the path. These values of y give a continuous path in the y-plane.

When x describes a closed path, say from x_0 back again to x_0, will the path of a selected point y be closed? The answer to this question depends on whether the change of θ' along the y-path is, or is not, a multiple of 2π; that is, on whether the change of θ (= twice the change of θ') is or is not a multiple of 4π; and this again on whether the path of x encircles the origin an even or an odd number of times.

When this number is even the y-path is closed. When, for example, x describes a circle which does not include $x = 0$, the change of θ, and therefore of θ', along this circle is 0, and the y-path is closed. But when the number is odd, the path of y is

not closed. If it began at y_0 with the amplitude θ' it ends with the same absolute value as y_0 but with the amplitude $\theta' + \pi$; *it has, in fact, become* $-y_0$ *and the values of* y *have been interchanged.* Here we meet with a phenomenon which can only present itself in the case of *many-valued* functions. Because $\sqrt{x}$ can be made to pass continuously into $-\sqrt{x}$, the two determinations of y are to be regarded not as separate functions but as *branches* of a two-valued function. We have already, in ch. XIII., had occasion to introduce the notion of a branch of a many-valued function. We shall now consider, in connexion with the present example, what is implied in this notion. (1) In ch. XIII. the separation of the functions into branches was determined in accordance with the requirement for the chief amplitude, namely $-\pi < \theta \leqq \pi$ (see, e.g., § 100). In the present case also we can give precision to our ideas by adopting the same rule; thus $x^{1/2}$, $-x^{1/2}$ become the two branches in question. Each varies continuously with x, except when x crosses the axis of negative real numbers. To express what happens in this case it is convenient to treat this axis as having two *banks*, a *right bank* and a *left bank*; the right bank is to be regarded as belonging to that half of the plane in which the coefficient of i in x is positive, and the left bank is assigned to the other half-plane. When x crosses from the right bank to the left (or conversely), $x^{1/2}$, $-x^{1/2}$ pass discontinuously into $-x^{1/2}$, $x^{1/2}$. The discontinuity of $x^{1/2}$ and $-x^{1/2}$ across this half-axis is caused by the discontinuity of the chief amplitude. (2) If we write $\sqrt{x} = \sqrt{c + (x - c)}$ and use the binomial theorem we get two power series in $x - c$. When the domains of these series do not contain negative real numbers, the two series belong wholly to $x^{1/2}$, $-x^{1/2}$ respectively: but when they do contain negative real numbers, each domain is divided into two parts by the axis of negative real numbers. The values taken by such a power series in the parts above and below this axis belong to different branches. (3) It is not essential that the separation of the two-valued function into branches should be made by means of the straight line from 0 to $-|\infty|$; any line from 0 to ∞ which does not intersect itself will serve our purpose.

Let such a straight line be treated as a cut in the x-plane over which no x-path is to be allowed to pass. With this restriction on the freedom of motion of x, we allow x to start from a point x_0 and describe all possible paths, and make the corresponding initial values of y, say $y_0, -y_0$, vary *continuously* along such paths; the result is two aggregates of values which constitute two *branches* of the two-valued function. Each of these two branches is, in general, composed partly of values $x^{1/2}$ and partly of values $-x^{1/2}$.

In the case of a many-valued analytic function there are as many branches as there are values of the function for an arbitrary value of x.

To take a special example let us see how two branches of $\sqrt{x}$ pass into each other when x starts from $x = 1$ and describes a unit circle about the origin; in this case the corresponding y-points start at $+1$ and -1 and describe a unit circle, but only half as fast. When x has completed its circle the y-points which began at ± 1 are at ∓ 1; together they describe the whole circle.

From what we have said it is clear that the origin has some speciality of position. The *analytic* peculiarity of the origin is that there are no power series Px for the function, though there are power series $P(x-c)$, when $c \neq 0$; we have to content ourselves with two terminating fractional series $x^{1/2}$, $-x^{1/2}$. A *geometric* peculiarity is that the isogonality breaks down; for, angles at $x = 0$ are not equal to but double of the corresponding angles at $y = 0$.

It is desirable to have a distinctive name for a point x_0 where a many-valued function of x has the double property that (1) two or more values become equal, and (2) the corresponding values permute after a complete description by x of a small circle whose centre is x_0. Such points are called *branch-points*.

In the present example $x = 0$ is a branch-point; so also is $x = \infty$. In justification of the latter part of this statement observe that ∞ is a single point in the theory of functions,—so that the two values of $\sqrt{x}$ are equal when $x = \infty$,—and further that a positive description of a circle which includes the origin

can be treated as a negative description of a circle which includes $x = \infty$, and therefore the branches permute round $x = \infty$.

The difficulty which arises from having two points in the y-plane corresponding to one in the x-plane was obviated by Riemann. He supposed the x-plane covered by two sheets, parallel and infinitely near to one another. Thus to a given point x correspond two *places*: one in the upper sheet, one in the lower. One of these places corresponds to one y-point, the other to the other. A place is named by its x and the corresponding y; thus, supposing the sheets horizontal, two places in the same vertical will be (x, y) and $(x, -y)$. We shall call them *co-vertical*. At the origin $x = 0$ we have only one place since the y's are equal; so we regard the sheets as stuck together (or in contact) at the origin. But a circle round the origin is to lead from the place (x, y) to the place $(x, -y)$. Therefore there must be a bridge between the sheets, which may run from 0 to ∞ along an arbitrary line, say along the ray of positive real numbers. Whenever a moving place crosses this bridge it passes from the upper to the lower or from the lower to the upper sheet. Along the bridge the surface intersects itself. Fig. 61 shows a vertical section of the surface, perpendicular to and intersecting the bridge; and fig. 60 gives an idea of the appearance of the surface as a whole*.

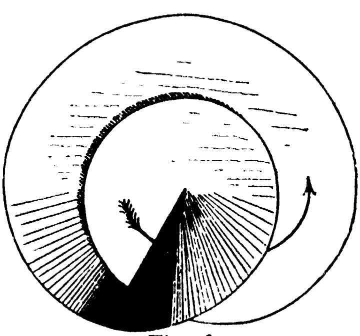


Fig. 60.

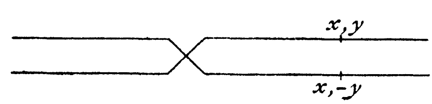


Fig. 61.

This Riemann surface answers all the requirements:—these are that two geometric points shall correspond to a given x and

* Fig. 60 is reproduced from a photograph taken from the model in Brill's collection.

that a path which passes round the origin an odd number of times shall lead from one y-point to the other, while a path which passes round the origin an even number of times shall lead from a y-point to itself. Thus on our surface a circle is a closed path where it does not include the origin; for if it cross the bridge once, say into the lower sheet; it recrosses it into the upper sheet. But a circle which includes the origin is no longer a closed path on the surface; it becomes closed when described twice. The y-plane and the x-surface are so related, point-to-point, that any closed path on the one corresponds necessarily to a closed path on the other.

What we have called (provisionally) the bridge will serve as a cut in the x-plane which determines two branches of the function; in this case the two branches are assigned to the upper and lower sheets respectively. When, conversely, a cut has been employed to create branches, it is often convenient to use that cut as a bridge on the Riemann surface, and to call it a *branch-cut*. For instance when we choose $x^{1/2}$, $-x^{1/2}$ for the two branches, the axis of real negative numbers is a branch-cut for the corresponding Riemann surface.

148. Corresponding Paths in the x-, y-planes when $y^2=x$. Before leaving this simple case let us determine some corresponding paths in the planes of x and y. We have seen that if ρ, θ and ρ', θ' be the polar coordinates of x and y, then

$$\rho=\rho'^2,\ \theta=2\theta'.$$

Hence the passage from a given polar equation of a y-curve to the polar equation of the corresponding x-curve, and vice versa, is immediate. Thus the circle $\rho'=\text{constant}$ maps into the repeated circle $\rho=\text{constant}$; two rays $\theta'=a$ and $\theta'=a+\pi$ into the repeated ray $\theta=2a$; the lines $\rho'\cos(\theta'-a)=\pm\kappa$ map into the repeated curve $\rho\cos^2(\theta/2-a)=\kappa^2$, which is a parabola.

Ex. Draw the maps in the x-plane of the lines $\xi'=0,\ \pm 1,\ \pm 2$, and $\eta'=0,\ \pm 1,\ \pm 2$, where $y=\xi'+i\eta'$.

The repeated line $\rho\cos(\theta-a)=\kappa$ maps into the curve $\rho'^2\cos(2\theta'-a)=\kappa$, which is a rectangular hyperbola with centre at $y=0$; and so on.

Ex. Draw the maps in the y-plane of the lines $\xi=0,\ \pm 1,\ \pm 2$, and $\eta=0,\ \pm 1,\ \pm 2$, where $x=\xi+i\eta$.

These examples show how we can identify, by means of the polar equation, the curve that corresponds to a given curve. But it should be observed that the mapping itself gives a clear idea of the form of the new curves. This process deserves some illustration.

Thus when $x=y^2$ we have $dx/2\sqrt{x}=dy$. Now when y describes a straight line am dy is a constant, say a. Hence the x-curve is such that

$$\text{am}\, dx - \tfrac{1}{2}\,\text{am}\, x = a.$$

That is, if ϕ be the amplitude of the tangent at a point, θ the amplitude of the stroke from $x=0$ to that point,

$$2\phi - \theta = 2a,$$

or
$$\phi - \theta = 2a - \phi\,;$$

and this expresses the well-known fact that rays issuing from the focus of a parabola are reflected at the curve so as to be parallel

So again when x describes a straight line through a point a, we have

$$\text{am}\,(x-a) = \text{constant}.$$

If $b^2=a$, then
$$y^2 - b^2 = x - a\,;$$

therefore, for the corresponding curve,

$$\text{am}\,(y^2-b^2) = \text{constant},$$

or
$$\text{am}\,(y-b) + \text{am}\,(y+b) = \text{constant}\,;$$

or the sum of the angles made with a given line by the lines from the ends of a given diameter to any point is constant. And this is a characteristic property of the rectangular hyperbola.

Next let the point x describe a circle, with centre a. Then, if $b^2=a$, we have

$$y^2 - b^2 = x - a,$$

and
$$|y-b|\,|y+b| = |x-a| = \text{a constant}.$$

Thus the product of the distances of any point of the y-curve from two fixed points b, $-b$ is constant. Such a curve is called a *cassinian*. Fig. 62

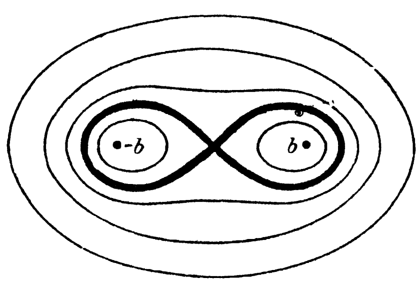


Fig. 62.

shows the map in the y-plane of a system of concentric x-circles. It can be shown that when the x-circle does not include the origin, the cassinian consists of two ovals ; when it does include the origin, the two ovals unite into a unifolium. If we use the Riemann surface we have in the former case two separated circles, one in each sheet, corresponding to the two ovals ; but in the latter case a single repeated circle corresponding to the unifolium.

In the separating case the cassinian becomes a figure-of-eight; this particular cassinian is known as the *lemniscate.*

The diameters of the concentric system of circles are given by $\operatorname{am}(x-a)=$ constant; these straight lines map into the curves

$$\operatorname{am}(y-b)+\operatorname{am}(y+b)=\text{constant}.$$

By the property of isogonality these curves cut at right angles the system of cassinians. It may be left to the reader to show that these curves are rectangular hyperbolas which have b, $-b$ for ends of a common diameter.

The above orthogonal curves admit of an easy generalization. Let any rational function $R(y)$ of y be put equal to x. Let x describe a circle, say about the origin as centre; then the corresponding path of y is given by

$$|R(y)|=\text{constant},$$

or

$$\rho_1\rho_2\rho_3\ldots\ldots=\kappa\rho'_1\rho'_2\rho'_3\ldots\ldots,$$

where the ρ's and ρ''s are the distances of y from the zeros and infinities of $R(y)$. The orthogonal system of curves will be the maps of straight lines through the origin, and will therefore have the equation

$$\theta_1+\theta_2+\theta_3+\ldots=a+\theta'_1+\theta'_2+\theta'_3+\ldots,$$

where $\theta_1, \theta_2, \ldots$ are the amplitudes of the strokes to y from the zeros, $\theta'_1, \theta'_2, \ldots$ are the amplitudes of the strokes to y from the infinities of $R(y)$, and a is a parametric constant. When y passes to ∞ the amplitudes of $y-a$ and $y-b$, where a and b are given, tend to become equal. Hence for the real points at ∞ on a curve of this system, we have

$$\theta_1=\theta_2=\ldots=\theta'_1=\theta'_2=\ldots=\beta \text{ say};$$

so that, if there are n zeros and n' infinities in the finite part of the plane, we have

$$(n-n')\,\beta\equiv a\,(\text{mod. } 2\pi),$$

showing that the curve has $n-n'$ asymptotes inclined at equal angles $2\pi/(n-n')$.

Ex. Prove that these asymptotes meet at a point.

Returning to the equation $y^2=x$, let us see what corresponds to a y-circle. Let the centre be b, and write $y=b+y'$, $x=b^2+x'$.

Then

$$x'=2by'+y'^2.$$

Regard y' as a point describing its circle with a constant angular velocity; then y'^2 describes a circle with double that angular velocity. The motion of x' arises then from a superposition of two circular motions; $2by'$ describes a circle with a certain constant angular velocity, and y'^2 describes a circle about $2by'$ with double that angular velocity. The composition of these two motions is a familiar question in Kinematics, and the resulting curve is called a *limaçon.*

This curve has different shapes according as the y-circle does or does not include the point $y=0$. When it does not, the curve is a unifolium (fig. 63, 1); but when it does, the curve subtends an angle 4π at the branch-

point $x=0$, winding twice round it. The curve in this case (fig. 63, 3) crosses itself at a point x_0 which can be determined as follows.

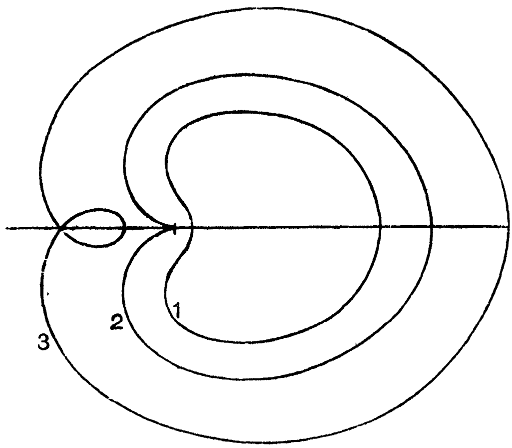


Fig. 63.

While y describes the above circle, say C, $-y$ describes another equal circle C'. On the *x-plane* these map into the same limaçon, for the points y, $-y$ map into a single point x on the x-plane; but on the *x-surface* they map into two limaçons L and L', a vertical cylinder through L cutting out on the surface the other limaçon L'; for the points y and $-y$ map into covertical places (x, y), $(x, -y)$ of the x-surface. When C and C' do not intersect, L and L' do not intersect; but when C and C' do intersect at points y_0, $-y_0$, then L and L' also intersect and the places of intersection are (x_0, y_0) and $(x_0, -y_0)$, where $x_0 = y_0{}^2$. Hence to find x_0 we take the map of either point of intersection of C and C'.

Ex. When C is the circle $(1, \rho)$, the point x_0 is $1-\rho^2$.

In the separating case when the circle passes through the point $y=0$, the limaçon has a cusp; for to the angle π at $y=0$ corresponds the angle 2π at $x=0$, so that when y passes through its origin, x reverses its direction. This special limaçon is called the *cardioid* (fig. 63, 2).

149. Example II. $y^2 = (x-a)/(x-b)$. The correspondence of x and y is 1, 2; accordingly we suppose (when the matter is to be discussed fully) a two-sheeted Riemann surface spread over the x-plane. The values of y are equal when $x=a$ and when $x=b$. The point a is a branch-point, for a circle about a which does not include b increases am$(x-a)$ by 2π and leaves am$(x-b)$ unaltered, so that the amplitude of a selected y is increased by π, not 2π, and one branch passes into the other; for a similar reason b is also a branch-point.

Let x_0, y_0 be a pair of values satisfying the equation. When x describes any path starting from and returning to x_0, either y returns to y_0, or the final value is $-y_0$. To distinguish we must examine the angle made by the path at both a and b.

When the whole changes of am $(x-a)$ and of am $(x-b)$ are $2\alpha\pi$, $2\beta\pi$, then that of y is $(\alpha-\beta)\pi$. The y-path is therefore closed when $\alpha-\beta$ is even, not closed when $\alpha-\beta$ is odd. The effect of a path depends *solely* on these angles $\alpha\pi$, $\beta\pi$; therefore we can continuously deform the path provided we do not let it cross a branch-point. The path can be resolved into a succession of circuits beginning and ending at x_0 and not intersecting themselves (fig. 64). A circuit which does not include a branch-point is one for which $\alpha=0$, $\beta=0$, and may be left out, since it has no effect on y_0. A circuit which includes both branch-points is one for which $\alpha=\beta=1$, and may also be omitted. The only circuits which affect the question are those which include a alone and b alone.

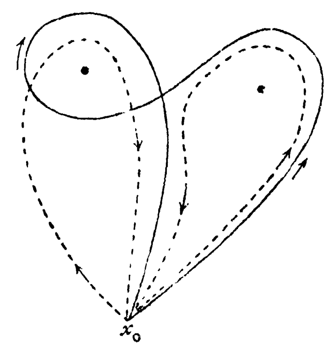


Fig 64

Now a bridge between the sheets from the branch-point a to the branch-point b will ensure that a moving place on the Riemann surface shall change from one sheet to another when it makes a circuit round a or b; that is, it will lead from the place (x_0, y_0) to the place $(x_0, -y_0)$. Thus by supposing such a bridge our Riemann surface is completed. The bridge may be along any curve provided its ends are a and b and it does not cut itself. Using this bridge as a branch-cut the two sheets determine two branches, as in § 147.

Observe that the example $y^2=x$ is closely connected with the present example. For if we write $z=(x-a)/(x-b)$, then

$$y^2=z$$

becomes $\quad y^2=(x-a)/(x-b),\ z=(x-a)/(x-b).$

The transformation $z=(x-a)/(x-b)$ is to be regarded as applied to *both sheets* of the surfaces spread over the x- and

z-planes; so that two covertical places of the x-surface become two covertical places of the z-surface. When the covertical places of the x-surface become one at a branch-place, so do those of the z-surface. Thus the branch-places a, b of the one surface map into the branch-places 0, ∞ of the other surface.

150. Example III. Rational Functions of x, y where $y^2 = (x-a)/(x-b)$. On the Riemann surface just described, y is a one-valued function; that is, for a given place on the surface y has one value only. In fact the *raison d'être* of a Riemann surface is to have one value of a dependent variable y at each point of *an x-surface*, instead of several values of y at each point of *an x-plane*.

But also a rational function $z, = R(x, y)$, of x and y has one value only, when x and y are both given, and therefore is one-valued on the surface. If the function be even in y it is of course one-valued in x alone, and the *x-plane* will suffice for our purposes. When, however, z contains odd powers of y, it has two values for a given x, differing only in the sign of the square root $\sqrt{(x-a)/(x-b)}$. These become equal when the square root is 0 or ∞; that is, when $x = a$ or b. Moreover they are interchanged when the square roots are interchanged. Thus the surface will serve for the representation of the pairs (x, z), equally with the pairs (x, y); but it must be observed that now the pairs (x, z), $(x, -z)$ do not occur at covertical places.

As an example let us consider the rational function

$$z = y(x-b).$$

Then $$z^2 = y^2(x-b)^2 = (x-a)(x-b).$$

We need no new surface; we have still the branch-places a, b, and the necessary bridge between a and b. But the relation of x to z is quite different from the relation of x to y. Whereas, corresponding to a given y there was one x, to a given z there are two x's. Thus in a complete discussion of the 2, 2 correspondence given by

$$z^2 = (x-a)(x-b),$$

we should be led to consider two Riemann surfaces:—a two-

sheeted surface spread over the x-plane and a two-sheeted surface spread over the z-plane. There would be a 1, 1 correspondence between the places (x, z) of the x-surface and the places (z, x) of the z-surface.

In this example we have, by eliminating y, passed from the simple to the more complex; and the reverse is the proper order. To consider the relation $z^2 = (x - a)(x - b)$, we can write $z = y(x - b)$, and then consider $y^2 = (x - a)/(x - b)$, so that x is a rational function of y, and therefore z also is a rational function of y.

The reader is familiar already with this order of ideas in connexion with integration, for the above is one of the ways, in elementary Integral Calculus, of reducing $\int\sqrt{(x-a)(x-b)}\,dx$ to the integral of a rational function. The transformation $\sqrt{(x-a)(x-b)} = y(x-b)$ leads to $\int 2(a-b)^2 \frac{y^2}{(1-y^2)^3}\,dy$.

The equation $z^2 = (x - a)(x - b)$ presents a new feature. As before two values of z are equal when $x = a, b$; but now, in addition, two values of z are equal when $x = \infty$. There is however no change of branches round $x = \infty$; for a path round both a and b adds 2π to each of the quantities $\operatorname{am}(x-a)$, $\operatorname{am}(x-b)$ and therefore 4π to $\operatorname{am} z$, thus restoring the initial z. Analytically we have

$$z = \pm x(1 - a/x)^{1/2}(1 - b/x)^{1/2};$$

whence expanding by the binomial theorem we have for large values of x, not a series containing fractional powers, but a Laurent series with one positive power, namely a term in x. This answers to the case of a *node* in the plane curve and will be called the *nodal case*. The sheets of the x-surface touch at $x = \infty$ and so do those of the z-surface at $z = \infty$; but there is no other connexion between the sheets in the neighbourhoods of those places. We return to this in § 158; before passing on observe a distinction between this phenomenon and the one which presented itself in the case of branch-points. Consider the two equations $(y - b)^2 = x - a$, and $(y - b)^2 = (x - a)^2 + (x - a)^3$. In the former case two values of y are equal to b when $x = a$, but only one value of x is equal to a when $y = b$. In the latter case

(the nodal case) not only are two values of y equal to b when $x=a$, but also conversely two values of x are equal to a when $y=b$.

151. Example IV. $y^3-3y=2x$. The correspondence of x and y is 1, 3; we require a three-sheeted x-surface. The values of y corresponding to a few values of x are shown by the table

$x=$	0	1	-1	∞
	0	-1	1	∞
$y=$	$3^{1/2}$	-1	1	∞
	$-3^{1/2}$	2	-2	∞

Here we are careful to include all cases where values of y become equal. These are sure to lead to branch-points because x is one-valued in y and therefore there are not two values of x to give rise to the nodal case

It simplifies the explanation to draw the curve which represents the correspondence of the real values of x and y, (fig. 65).

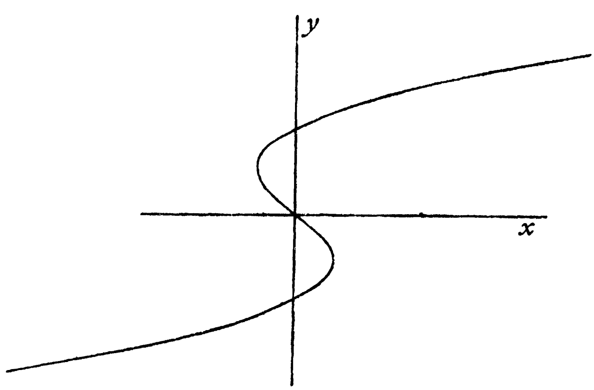


Fig. 65.

Taking the three sheets, we assign the values $3^{1/2}$, 0, $-3^{1/2}$ to the three places $x=0$ in any order, say $3^{1/2}$ to the upper sheet, 0 to

the middle one, $-3^{1/2}$ to the lower sheet. When a place s on the surface starting from $(0, 0)$ moves to the right along the real axis, a glance at the curve shows that the corresponding value of y decreases from 0 to -1; while if the place start from $(0, -3^{1/2})$ and move in the same way, the values of y increase from $-3^{1/2}$ to -1. Thus the middle and lower sheets unite at $x = 1$; on the upper sheet $(1, 2)$ is an ordinary place.

So if we allow x to pass through real values from 0 to -1, the curve shows that the places which unite at $x = -1$ are those which start from $(0, 0)$ and $(0, 3^{1/2})$. Thus the upper and middle sheets unite at $x = -1$.

We draw a branch-cut from each branch-place to ∞ connecting the sheets which unite at that branch-place; the two branch-cuts are arbitrary in form, but must not intersect themselves or one another. We will suppose them drawn along the real axis (fig. 66). In the figure the paths in the three sheets are represented respectively by a continuous line, a line with dots, and a broken line.

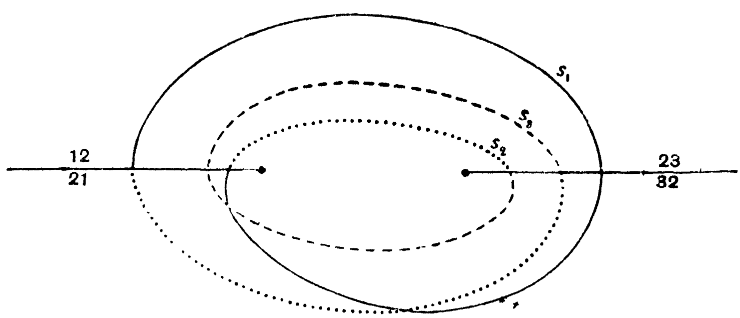


Fig. 66.

The construction of the surface is now effected. To each place on it corresponds a definite value not only of x but also of y.

Notice how the requirements at ∞ are satisfied. When x is large the values of y are approximately $(2x)^{1/3}$, $v(2x)^{1/3}$, $v^2(2x)^{1/3}$, where $v = \exp(2\pi i/3)$; thus a large circle round ∞ is to interchange these three values *cyclically*. Now a place s which starts

at a place s_1 in the upper sheet and north of the real axis and describes about the origin a large circle positively, will pass successively into the middle and lower sheets before reaching a place s_3 in the same vertical with the initial place. That is, it returns after one revolution to a place s_3 in the lower sheet. In another revolution it passes under the branch-cut (12) and passes at the branch-cut (23) into the second sheet; thus after the middle revolution it is at s_2 in the middle sheet, where s_2 is vertically below s_1. A third revolution brings it back to s_1 and its path is closed.

Ex. Draw on the x-plane and also on the x-surface the maps of the circles $|y|=1, |y|=2$.

152. Simply connected Riemann Surface. Starting with a given equation $R(y)=x$, where $R(y)$ is a rational function of degree n in y, we have for the adequate geometric representation a y-plane and an n-sheeted x-surface with certain branch-places; of course the nodal case does not occur since there are not two values of x to become equal. These two surfaces are in place-to-place correspondence. On the x-surface y is one-valued and so is any arbitrary rational function z of x and y. Eliminating x we have z as a rational function of y, and we can by means of this relation map the y-plane on a z-surface.

The places of the x-surface and the z-surface are in 1, 1 correspondence with those of the y-plane and therefore with one another. There are of course, according to the assumed rational function z, different z-surfaces; but all have one property in common. This we proceed to explain.

If we draw a circuit on the y-plane we divide the plane into two regions,—an inside and an outside,—such that it is not possible to pass from the one region to the other without crossing the circuit. To this circuit corresponds a circuit on any z-surface which must also divide that surface in the same way. For on account of the correspondence a path which crosses the circuit on the plane maps into a path which crosses the corresponding circuit at the corresponding place on the surface.

If the plane be cut along the circuit it separates into two

parts; if then the surface be cut along the corresponding circuit it also falls apart. This is a fact characteristic of surfaces which can be mapped with 1, 1 correspondence on a plane. For it can be proved that only surfaces which fall apart when cut along *any* circuit can be so mapped. Such surfaces when cut along any circuit form two simply connected surfaces (§ 107).

Given the possibility of mapping there still remains the further and important question as to how it is to be done.

The next example will yield a surface which is not simply connected.

153. Example V. $y^2 = (x-a_1)(x-a_2)(x-a_3)(x-a_4)$. Here we require a two-sheeted x-surface with branch-places at a_1, a_2, a_3, a_4. When $x=\infty$ the values of y are equal, but a large circuit round a_1, a_2, a_3, a_4 increases am y by 4π, so that there is no interchange of values round ∞; that is, ∞ is a nodal place.

When x describes in the *x-plane* positively (or negatively) a path C which starts from x_0 (say) and passes once round one of the points a, say a_2, but does not include any other, the two values

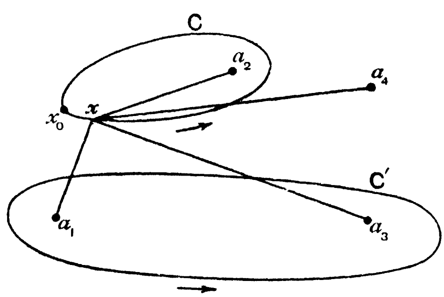


Fig. 67.

of y permute (fig. 67); for am $(x-a)$ is unaltered for $a=a_1, a_3, a_4$, and increases by 2π for $a=a_2$. When x describes a path C' round two branch-points, say a_1, a_3, similar reasoning shows that the two values of y are restored at the end of the path. It is easy and useful to construct paths that are more complicated and to take account of their effects on the two values of y.

All these effects are provided for by the construction of two branch-cuts on the Riemann surface, from a_1 to a_2 and from a_3 to a_4 respectively.

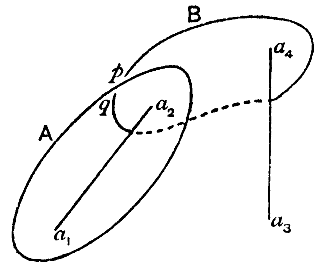


Fig. 68.

There is an essential distinction between this surface and those previously considered. The present surface is not simply connected. Draw a circuit A (fig. 68) in one sheet round the bridge a_1a_2. Then the path B leads from a place p on the one side of A to a place q on the other without crossing A. Thus the surface if cut along A will not fall apart; it is still one surface bounded by the two rims of the cut; having been to begin with a surface with no boundaries at all. If the cut surface be cut again from p to q along B it still does not fall apart, for it has *one* continuous boundary as shown in fig. 69.

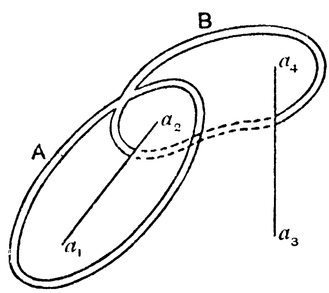


Fig. 69.

But in point of fact any further cut from the boundary to the boundary, or any cut along a closed circuit, will sever the surface into two surfaces; so that when cut along A and B the region is simply connected. And (assuming this) it can then be shown that any circuit on the surface is deformable into repetitions of the circuits A and B.

This discussion of the connectivity of a Riemann surface is fundamental in Riemann's theory of the integrals of algebraic functions, or Abelian integrals as they are called in general. There are two special cases which might find a place here; namely the integrals $\int R(x, y)\,dx$, where R stands for a rational function of its arguments and (1) y is a rational function of x, (2) y is given in terms of x by the equation of Ex. V.; these cases are often called the *rational* and *elliptic* cases, because they bring in (1) integrals of rational functions of x, (2) elliptic integrals. A discussion of these two special cases would illustrate

the advantages of Riemann's methods; but we shall not undertake any such discussion here, as the methods referred to belong properly to the whole theory of Abelian integrals and not merely to these special cases.

154. Fundamental Regions. A system of operations is said to be a *group* when every operation compounded of any number of operations taken from the assigned system is itself an operation of the system. For example, translations in the x-plane form a group, for a translation combined with a translation is a translation; again the bilinear substitutions

$$x' = (ax+b)/(cx+d) \text{ or } (x, (ax+b)/(cx+d))$$

form a group, for if x', x and x'', x' are connected by bilinear relations, so also are x'', x. We can impose in this latter case the restriction that a, b, c, d are to be *integers* which satisfy the equation $ad-bc=1$ and still we get a group.

To take a much simpler example, the bilinear substitutions

$$x' = e^{2r\pi i/n}x \qquad r = 0, 1, 2, \ldots, n-1$$

form a group. In this group every substitution can be expressed as a power of the substitution $S=(x, e^{2\pi i/n}x)$; in fact the group is $1, S, S^2, \ldots, S^{n-1}$, where $S^n=1$, 1 standing for (x, x). Such a group is called a *cyclic* group and S is called the generating substitution of the group. The substitution S is a special elliptic substitution (§ 37) for which the fixed points are $0, \infty$; it arises from the substitution

$$\frac{x'-x_0}{x'-x_1} = e^{2\pi i/n}\,\frac{x-x_0}{x-x_1}.$$

In place of the factor $e^{2\pi i/n}$ we may take $e^{i\theta}$ and still have a cyclic group; when θ/π is not a rational number the group contains infinitely many distinct substitutions S^r, and there is no longer a relation $S^n=1$.

That there may be important connexions between groups and functions is indicated by what we know of the elementary functions x^n, e^x and of the elliptic function $\wp u$. The first of these is invariant with respect to the group of substitutions $(x, e^{2r\pi i/n}x)$, the second with respect to the group of substitutions $(x, x+2r\pi i)$,

and the third with respect to the group of substitutions $(u, \pm u+w)$ where $w=2m_1\omega_1+2m_2\omega_2$, m_1 and m_2 being integers. Given any one-valued analytic function fx and all the bilinear substitutions $(x, (ax+b)/(cx+d))$ which do not affect the value of the function, it is clear that we get a group, for the combination of two substitutions which do not affect fx is a substitution of the same character. Conversely it is important to connect, if possible, with an assigned group of bilinear substitutions a one-valued analytic function fx such that

$$f\left(\frac{ax+b}{cx+d}\right)=fx,$$

when $\left(x, \dfrac{ax+b}{cx+d}\right)$ is a member of the group. The discussion of this problem belongs to the subject of automorphic functions and lies beyond the range of this book; but we can gain from the special examples x^n, e^x, $\wp u$ a fairly good insight into the meaning of what are known as fundamental regions and a perception of their importance in the study of functions.

I. *Equivalent points.* Taking any point x of the plane and applying to it all the substitutions of a group, we get a set of *equivalent* points. For example the substitutions

$$(u,\ u+2m_1\omega_1+2m_2\omega_2),\quad \text{or}\quad (u,\ u+w),$$

form a group, and the points $u+w$ are equivalent points.

II. *Fundamental region of a group of bilinear substitutions.* In the case of x^n we divide the plane into n regions by the rays $\theta=-\pi/n, \pi/n, 3\pi/n, \ldots$; each region is to contain one but not both rays, for example the first region $-\pi/n$ to π/n contains the ray π/n, the second region π/n to $3\pi/n$ contains the ray $3\pi/n$ and so on. Select one of these regions, say the first; we shall call this region a *fundamental region of the group* because it contains one point and not more than one point which is equivalent with respect to the group of substitutions $(x, e^{2r\pi i/n}x)$ to an arbitrarily selected point x of the x-plane.

In the cases of e^x, $\wp u$ we are led to consider bands and parallelograms. With respect to the group of substitutions

$(x,\ x+2r\pi i)$, bands (§ 100) of breadth $2\pi i$ are fundamental regions; with respect to the groups of substitutions $(u,\ u+w)$ and $(u,\ \pm u+w)$ the fundamental regions are the parallelogram of periods and half the parallelogram of periods respectively.

III. *One-valued functions associated with fundamental regions.* When we take the substitution $(x,\ e^{2\pi i/n}x)$ which converts the ray $-\pi/n$ into π/n, and apply it successively to the fundamental region ($-\pi/n$ to π/n) we get in the x-plane n sheets laid side by side; the plane is covered once without gaps. Now let us consider the equation $y=x^n$; instead of spreading n sheets over the y-plane we can equally well take these n regions lying side by side in the x-plane. That there are n values of x for a given y can be inferred from knowing that x^n takes all its values in each of the regions: that the function x^n which passes once through all its values in the fundamental region is one-valued is a consequence of the non-overlapping of the aggregate of regions, and so on.

Suppose that we had started with the region ($-\pi/n$ to π/n) we could have evolved the group from it by observing that the substitution $(x,\ e^{2\pi i/n}x)$ converts the one edge into the other, and therefore will, on successive applications, rotate the initial region into the remaining regions. Also we could have inferred that the one-valued function (if any) which takes all its values once within the region must satisfy the relation $f(e^{2\pi i/n}x)=fx$, otherwise it would not have the same set of values along the two rays $-\pi/n$, π/n.

If we begin with a band bounded by two parallel straight or curved lines which are coordinated by the substitution $(x, x+\omega)$, we are led to the theory of one-valued functions fx which satisfy the relation $f(x+\omega)=fx$, and take their values once and only once in the fundamental region belonging to the group of substitutions $(x,\ x+r\omega)$. The infinitely many bands for e^x cover the plane once and without overlapping; this indicates that when $y=e^x$, y is one-valued and defined for all values of x, and that x is an infinitely many-valued function of y. Instead of the infinitely many bands in the x-plane we can use an

infinitely many-sheeted surface in the y-plane. When y is not allowed to cross the negative half of the axis of real numbers, x is restricted to a band; if we allow x to move into adjoining bands we must let y cross the barrier. To the complete x-plane then corresponds the infinitely many-sheeted y-surface with all the sheets hanging together at 0 and ∞ and connected by a branch-cut which replaces the barrier. The infinitely many bands in the x-plane serve equally with the infinitely many sheets of the surface spread over the y-plane to indicate a suitable separation of $\log x$ into branches.

Since $\sin x = \cos(\pi/2 - x)$ the fundamental region for $\cos x$ will serve also for $\sin x$. Now $\cos x$ is unaltered by the substitutions

$$(x, \ \pm x + 2r\pi);$$

hence the fundamental region is that given by this group of substitutions and not the band of breadth 2π given by the group of substitutions $(x, x + 2r\pi)$. Half, not the whole, of the latter band must be taken (§ 100). A similar remark applies to $\wp u$; here the group is $(u, \ \pm u + w)$, and hence a fundamental region for $\wp u$ is half of the cell

$$u_0, \ u_0 + 2\omega_1, \ u_0 + 2\omega_1 + 2\omega_2, \ u_0 + 2\omega_2.$$

The fundamental regions which we have been considering can be deformed in many ways. It will be sufficient if we point out that the dotted curvilinear parallelogram of fig. 70 can replace the rectilinear parallelogram without affecting the properties of the associated elliptic function $\wp u$.

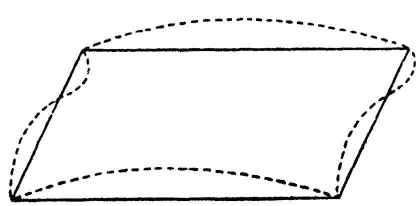


Fig. 70.

CHAPTER XXI.

ALGEBRAIC FUNCTIONS.

155 The Algebraic Function. Let $F(x, y)$ be a polynomial formed with positive integral powers of x and y; we suppose this polynomial to be *irreducible,* which means that it cannot be decomposed into the product of several factors of a similar kind but of lower degrees in the variables. The function y defined by the equation $F(x, y) = 0$ is called an *algebraic function* of x. Evidently x must also be called an algebraic function of y. When $F(x, y)$ is reducible, the equation $F = 0$ defines as many distinct algebraic functions as $F(x, y)$ admits distinct irreducible factors.

Let m, n be the highest powers of x and y. Then for a given x there are in general n finite values of y (§ 59); let us agree that there shall be in all cases n values of y, finite or infinite. Similarly when y is given we shall suppose that there are in all cases m values of x.

Thus the equation determines a correspondence of the x- and y-planes such that to an x-point correspond n y-points, and to a y-point m x-points.

There is an important difference between the usual point of view of projective geometry and that of the theory of functions. In the former the straight line is fundamental; because a certain straight line meets a conic in two points we arrange that all straight lines shali meet the conic in two points. For example the curve $xy = 1$ is met by the line $x = 1$ not merely

in $y=1$ but also in $y=\infty$. In the theory of functions on the other hand it is the correspondence of x and y that is insisted on; the equation $xy=1$ is a 1, 1 correspondence between x and y, and therefore to $x=1$ corresponds just one value $y=1$.

The algebraic equation $F(x, y)=0$ may be written

$$y^n f_0 x + y^{n-1} f_1 x + \dots + y f_{n-1} x + f_n x = 0 \text{ ......(1)},$$

where $f_0, f_1, \dots, f_n$ are rational integral functions of x of degrees not greater than m. The points at which a value of y becomes infinite are given by $f_0 x=0$. This equation has m roots; for to $y=\infty$ are to correspond m values of x. If then it is only of degree m_1 it has $m-m_1$ infinite roots.

It may happen that a value a of x which satisfies $f_0 x=0$ will also make $f_1 x=0$, $f_2 x \neq 0$. In this case the point ∞ figures twice among the points in the y-plane which correspond to $x=a$. If $f_0 a=0$, $f_1 a=0$, $f_2 a=0$, $f_3 a \neq 0$, there are three infinite values of y; and so on for higher cases.

Ex. If $y^2x+yx+y-1=0$, y is a two-valued function of x. For what values of x are the values of y equal? What is x when y is ∞? And what is y when x is ∞?

156. Proof that an Algebraic Function is Analytic. The theory of functions rests very largely on the basis of power series. The transition, then, from the above definition of an algebraic function to the series for y in powers of x is an essential preliminary to the knowledge of this function.

Suppose for a given finite x, say x_0, we have found a finite y, say y_0; and that the other values of y are all distinct from y_0. Writing $x-x_0$ for a new x and $y-y_0$ for a new y we can rearrange the equation (1) in the form

$$y = c_{10}x + c_{20}x^2 + c_{11}xy + c_{02}y^2 + \dots + c_{mn}x^m y^n \text{ ......(2)};$$

for by hypothesis when x is 0 there is to be one and only one y equal to 0, so that in the terms which contain y alone, y itself must occur. Let us consider more generally that the series on the right is infinite; and let no coefficient be greater than a given positive number γ, so that the infinite double series is convergent when $|x|<1$, $|y|<1$.

Then we can prove that there is one and only one power series in x without a constant term,

$$P_1 x = a_1 x + a_2 x^2 + a_3 x^3 + \dots \qquad\dots\dots\dots\dots\dots (3),$$

which, when substituted for y in the above equation (2), satisfies it identically; and that this series converges within an assignable region. For first if we assume that there is a series for y of the form (3), convergent when $|x| < R$, we can rearrange (2) in powers of x, for then the conditions of § 81 are satisfied. If then the series is to satisfy the equation (2) for all values of x such that $|x| < R$, we have, by equating coefficients, the equations

$$\left.\begin{array}{l} a_1 = c_{01}, \\ a_2 = c_{20} + c_{11} a_1 + c_{02} a_1^2, \\ \dots\dots\dots\dots\dots\dots\dots \end{array}\right\} \dots\dots\dots\dots\dots\dots (4),$$

by which in general a_n is given as a rational integral function of $a_1, a_2, \dots, a_{n-1}$. By these equations $a_1, a_2, \dots$ are uniquely determined in terms of the given c's, so that there can only be *one* such series as (3). The question is whether this series converges for values of x less than R where R is not zero. To settle this, replace all the c's by the positive number γ, such that γ is not less than the absolute value of any of the c's, and replace the system (4) by

$$\left.\begin{array}{l} \alpha_1 = \gamma, \\ \alpha_2 = \gamma + \gamma\alpha_1 + \gamma\alpha_1^2, \\ \dots\dots\dots\dots\dots\dots \end{array}\right\} \dots\dots\dots\dots\dots\dots (5),$$

that is, let $a_1, a_2, \dots$ become $\alpha_1, \alpha_2, \dots$ when for every c we write γ. Then evidently the numbers $\alpha_1, \alpha_2, \dots$ are all positive; and α_n is not less than $|a_n|$, for $|a_1| = |c_{01}| \leqq \gamma \leqq \alpha_1$,

$$\begin{aligned} |a_2| &= |c_{20} + c_{11} a_1 + c_{02} a_1^2| \\ &\leqq |c_{20}| + |c_{11} a_1| + |c_{02} a_1^2| \\ &\leqq \gamma + \gamma\alpha_1 + \gamma\alpha_1^2 \\ &\leqq \alpha_2, \end{aligned}$$

and so on.

Therefore the series (3) converges in a circle (R) if

$$\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \dots$$

does. Thus the question is reduced to this: can we assert that a power series $P_1 x$ for y exists when

$$y = \gamma x + \gamma x^2 + \gamma xy + \gamma y^2 + \ldots + \gamma x^m y^n + \ldots ?$$

But this is easily answered, for when $|x| < 1$ and $|y| < 1$, we can sum the series on the right and get

$$y = \gamma/(1-x)(1-y) - \gamma - \gamma y,$$

that is, $$(\gamma+1)y^2 - y + \gamma x/(1-x) = 0,$$

whence $$2(\gamma+1)y = 1 \pm \sqrt{\frac{1-(1+2\gamma)^2 x}{1-x}};$$

and the square root can be expanded in a power series so long as $|x| < 1$ and $|x| < 1/(1+2\gamma)^2$, the latter condition of course rendering the former unnecessary.

Thus we have obtained a circle of radius R within which the series (3) converges, namely the circle whose radius is

$$R = 1/(1+2\gamma)^2.$$

This being true of the infinite double series is *a fortiori* true on the supposition that all the c's after $c_{m,n}$ are zero, that is, for the finite double series (2). Then γ may be the greatest of the c's.

157. Puiseux Series. When $a_1 = 0$, $a_2 \neq 0$ then two values of x become equal to a when $y = b$. Let us now revert the series. We have

$$y - b = a_2(x-a)^2 + a_3(x-a)^3 + \ldots.$$

Taking square roots we have

$$\pm\sqrt{(y-b)/a_2} = (x-a)\sqrt{1 + P_1(x-a)}$$
$$= (x-a)P(x-a);$$

and now inverting the series we have, by § 84,

$$x - a = \pm\sqrt{(y-b)/a_2}\,[1 + P_1(\pm\sqrt{(y-b)/a_2})].$$

According to the square root selected we have then two distinct fractional series. These series are characteristic of a point b for which two values of the function x become equal. And in the same way if $a_1 = a_2 = \ldots = a_{r-1} = 0$, $a_r \neq 0$, we have

$$\sqrt[r]{(y-b)/a_r} = (x-a)[1 + P_1(x-a)],$$

where on the left we have a choice of r rth roots; whence, by reversion we have for $x - a$ a power series in $\sqrt[r]{(y-b)/a_r}$, and

thus r series from which to determine the behaviour of those r values of x which become equal to a when y is b.

Such fractional series do not of course occur for one-valued functions. They were employed with much effect by Puiseux in a classical memoir on algebraic functions; for shortness we shall call these fractional power series *Puiseux series*.

We have considered the Puiseux series in fractional powers of $y-b$ arising from nearly coincident values of x; we can of course interchange x and y and thus arrive at Puiseux series in fractional powers of $x-a$ arising from nearly coincident values of y.

When a series in $x-a$ associated with a root y is a Puiseux series, the point $x=a$ is called a *branch-point* (§ 147).

Referring to the simple case $y^2=x$, or $y=\sqrt{x}$, we saw that a circle about $x=0$ permuted the values of y. And what happened in this simple case happens generally. If by $\sqrt{x-a}$ we mean a selected square root, then a series

$$y-b=P_1(\sqrt{x-a})$$

is accompanied by another

$$y'-b=P_1(-\sqrt{x-a}),$$

which differs from the former only as regards the sign of the square root. When x describes a circle about a,—this circle lying in the region of convergence of the series,—then the square root changes sign and y becomes y' while y' becomes y.

And so generally if by $\sqrt[r]{x-a}$ we mean a selected root, a series

$$y-b=P_1\sqrt[r]{x-a}$$

is accompanied by $r-1$ other such series differing only in having another rth root; denoting $\exp(2\pi i/r)$ by α, the set of r series have no constant terms and are

$$\left\{\begin{aligned} y-b&=P\sqrt[r]{x-a},\\ y'-b&=P[\alpha\sqrt[r]{x-a}],\\ y''-b&=P[\alpha^2\sqrt[r]{x-a}],\\ &\;\;\vdots\qquad\qquad\vdots\\ y^{(r-1)}-b&=P[\alpha^{r-1}\sqrt[r]{x-a}].\end{aligned}\right.$$

When x describes a small circle about a, the amplitude of $x-a$ increases by 2π and therefore that of $\sqrt[r]{x-a}$ by $2\pi/r$. Thus $\sqrt[r]{x-a}$ becomes $\alpha\sqrt[r]{x-a}$ and so on; and this means that, *when x describes the small circle round a as centre, the values $y, y', y'', \ldots, y^{(r-1)}$ permute cyclically.*

Ex. The function $y=\sqrt[3]{x}+\sqrt{x}$ is six-valued. Mark the six points for which $x=1$ and follow their changes when x describes positively a circle about $x=0$ returning to the point 1.

158. Double Points on the Curve $F(x, y)=0$. So far we have considered (1) the ordinary case where a single value b corresponds to a single value a and conversely; (2) a special case in which r equal values b correspond to a single value a, while conversely to this value b corresponds only one value a; here

$$x-a=P_r(y-b),$$
$$y-b=P\left(\sqrt[r]{x-a}\right),$$

where the notation P_r indicates, as in § 75, that the series begins with $(y-b)^r$.

There remains for consideration another special case: that namely in which each of a selected pair of values a, b occurs more than once:—that is, when $x=a$, y takes r equal values b, and when $y=b$, x takes s equal values a. This question is too difficult to be treated here in all its generality; but we shall discuss the case $r=s=2$.

Taking a, b as new origins, then when y is zero x^2 is to be a factor and when x is zero y^2 is to be a factor; so that the algebraic equation is

$$c_{20}x^2+c_{11}xy+c_{02}y^2+\text{higher powers of } x, y=0.$$

We suppose here $c_{20}\neq 0$, $c_{02}\neq 0$; as otherwise more than two values of x (or y) will be zero when y (or x) is zero. Thus our equation is

$$(y-\alpha x)(y-\beta x)+\text{higher powers}=0,$$

where neither α nor β is 0. Notice that the two values of y which become equal to 0 when x is 0 satisfy not only $F=0$, $\partial F/\partial y=0$, but also $\partial F/\partial x=0$.

Let $y - \alpha x = zx$; then on substituting $(\alpha + z)x$ for y the equation becomes

$$z(\alpha - \beta + z) + xP(x, z) = 0,$$

where $P(x, z)$ is a terminating series of positive integral powers of x and z. Hence if $\alpha \neq \beta$, we have (§ 156)

$$z = P_r x \text{ where } r > 0,$$

and $$y = \alpha x + P_{r+1} x,$$

where the suffixes indicate that the series begin with the rth, $(r+1)$th terms respectively.

Similarly the factor $y - \beta x$ contributes a series

$$y = \beta x + P_{r'+1} x.$$

Hence we have power series in x for the two values of y which vanish when $x = 0$, *and $x = 0$ is not a branch-point although two values of y are equal when $x = 0$.*

But if $\alpha = \beta$ we have

$$z^2 + xP(x, z) = 0,$$

and it makes a difference whether or not $P(x, z)$ contains a constant term. If it does, then

$$x = P_2(z),$$

$$z = P\sqrt{x},$$

$$y = \alpha x + xP\sqrt{x} = P_2\sqrt{x},$$

and $x = 0$ is a branch-point; and similarly $y = 0$ is a branch-point.

The case in which $P(x, z)$ does not contain a constant term, —that is, when the terms of the second and third orders in the original equation have a common factor,—belongs to the more general question to which reference was made above; this is discussed in our larger treatise (ch. IV.).

159. Infinite Values of the Variables. There remains for consideration the case in which one or both of the selected pair of values a, b are ∞.

First when a alone is ∞, we determine the series for $y - b$ by writing $x = 1/x'$; points near ∞ in the x-plane become points

near 0 in the x'-plane and we can determine from the algebraic equation between y and x', *in the normal case*, a series

$$y - b = x'P_0 x' = \frac{1}{x} P_0\left(\frac{1}{x}\right)$$

for $y-b$. This is equivalent to replacing $x-\infty$ by $1/x$. Special cases can be treated as before.

Similarly when b alone is ∞ we write $y = 1/y'$ and discuss the algebraic equation between y' and x for the pair of values $x = a$, $y' = 0$.

Lastly when both a and b are ∞ we write $x = 1/x'$, $y = 1/y'$ and discuss the algebraic equation between x' and y' for the values $x' = 0$, $y' = 0$.

If for example $F(x, y) = 0$ take the form

$$y^2 x - x^2 - y^2 - 1 = 0,$$

then when $x = \infty$ both values of y are ∞, and when $y = \infty$ either $x = \infty$ or $x = 1$.

Writing $x = 1/x'$, $y = 1/y'$, the equation becomes

$$x' - y'^2 - x'^2 - x'^2 y'^2 = 0;$$

whence $$x' = P_2 y' \text{ and } y' = P_1(\sqrt{x'}).$$

The actual coefficients in the series are readily determinable in this simple case, for we can solve the equation at once for either x' or y'.

Near $y = \infty$, $x = 1$; we have, writing $y = 1/y'$, $x - 1 = x'$,

$$1 + x' - y'^2(1 + x')^2 - 1 - y'^2 = 0,$$

or $$x' - 2y'^2 - 2y'^2 x' - y'^2 x'^2 = 0;$$

whence $$x' = P_1 y'^2,$$

or $$x - 1 = P_1(1/y^2),$$

and $$1/y = P_1(\sqrt{x-1}),$$

so that

$$y = \frac{c_{-1}}{\sqrt{x-1}} + c_0 + c_1\sqrt{x-1} + c_2(\sqrt{x-1})^2 + \dots + c_n(\sqrt{x-1})^n + \dots.$$

This is a case in which an *infinity* of y,—that is, a value of x

which makes y infinite,—leads to a Puiseux series; in such a case the infinity is also a branch-point.

160. The Singular Points of an Algebraic Function. The general expression for the values of y which are infinite when $x=a$ is obtained from

$$1/y = P_r \sqrt[s]{x-a},$$

giving
$$y = \sqrt[s]{(x-a)^{-r}}\, P_0 \sqrt[s]{x-a}.$$

A function which is expressed near $x=a$ by a series of this form is said to have *an algebraic infinity* at $x=a$; it is essential to the existence of such an infinity that the number of terms with negative exponents shall be finite.

The singular points a of an algebraic function fx are (1.) points at which one or more values of fx become infinite, but in such a way that the corresponding expansions of $1/fx$ are of the form $P(x-a)$, or $P(1/x)$ if $a=\infty$; (2) branch-points near which several values of y are expressible in the form

$$\sqrt[s]{(x-a)^r}\, P_0(\sqrt[s]{x-a});$$

here r may equal zero or any integer (positive or negative), and $x-a$ is to be replaced by $1/x$ if $a=\infty$. The expansions at singular points of the kind (1) are of the same form as that at a non-essential singular point of a one-valued function.

An algebraic function has, then, the following properties:—

I. *Near a point $x=a$ it has, in general, n distinct finite values which are given by series $y-b_r=(x-a)\,P(x-a)$;*

II. *At a finite number of exceptional points some of the n series for y in terms of $x-a$ have to be replaced by series*

$$\sqrt[s]{(x-a)^{-r}}\, P_0 \sqrt[s]{(x-a)},$$

where $s=1$ and r is a positive integer, or $s=2, 3, \ldots$ and r is zero or any positive or negative integer.

The converse theorem that *an n-valued analytic function of x which has these two properties* I. *and* II. *is necessarily an algebraic function* is of great importance.

The proof results from the following considerations. Let

$y_1, y_2, \ldots, y_n$ be the n values of y; every symmetric function of these n values is one-valued in x. For example suppose that

$$y_1 = a_0 + a_1 \ (x-a)^{1/3} + a_2 \ (x-a)^{2/3} + \ldots,$$
$$y_2 = a_0 + a_1 v\,(x-a)^{1/3} + a_2 v^2\,(x-a)^{2/3} + \ldots,$$
$$y_3 = a_0 + a_1 v^2\,(x-a)^{1/3} + a_2 v^4\,(x-a)^{2/3} + \ldots,$$
$$y_4, y_5, \ldots, y_n = \text{power series in } x-a;$$

then

$$\sum_1^n y_r, \quad \sum_1^n y_r^2, \quad \sum_{r=1,\,s=2}^{n-1,\,n} y_r y_s \ (s>r),$$

etc. have no terms in $(x-a)^{1/3}$, $(x-a)^{2/3}$, $(x-a)^{4/3}$, ..., since the coefficient is in all cases $1+v+v^2$ and this vanishes. As there are no fractional exponents the corresponding series are of the form $P(x-a)$. The argument applies generally and shows that the symmetric functions of $y_1, y_2, \ldots, y_n$ are equal to power series in $x-a$, preceded it may be by a *finite* number of powers $(x-a)^{-r}$ for which the exponent is a negative integer, and a similar result holds for the other branch-points (which, it will be recalled, are finite in number). Hence all symmetric functions of $y_1, y_2, \ldots, y_n$ are one-valued functions of x with no singularities other than a finite number of non-essential singular points (∞ inclusive). It follows from the theorem of § 104 that these symmetric functions are rational functions of x; and as the coefficients $p_1, p_2, \ldots, p_n$ in

$$y^n + p_1 y^{n-1} + p_2 y^{n-2} + \ldots + p_n \equiv (y-y_1)(y-y_2)\ldots(y-y_n),$$

are equal to $-\Sigma y_1$, $+\Sigma y_1 y_2$, ..., $(-)^n y_1 y_2 \ldots y_n$, y satisfies an algebraic equation of degree n in y whose coefficients are rational functions of x.

The theorem that we have proved is a good illustration of the remark in § 104 that the character of a function is often determined best by observing the behaviour of the function at its singular points.

161. An Algebraic Equation in x, y defines a Single Function. We shall now show that it is possible to start from any non-singular point x_0 with any one of the n values of y at that point and by describing a suitable path arrive at an

arbitrary non-singular point x_0' with any one of the n values $y_1, y_2, \ldots, y_n$ that satisfy the equation in y for this value of x_0'.

To fix ideas let us start with a definite series

$$y - y_0 = (x - x_0) P (x - x_0)$$

at x_0; this can be continued so that one value of y shall be determined for any value x_0' so long as the path employed does not pass through a branch-point (and we shall suppose this to be the case for all paths employed). Suppose that one of the values at x_0', say $y - y_0' = (x - x_0') P (x - x_0')$, can by no means be attained; then starting from x_0' with this value and passing along any path to x_0 the value of y at x_0, say $Q(x - x_0)$, must be distinct from $y_0 + (x - x_0) P (x - x_0)$. This means that when x describes any circuit the values $y_1, y_2, \ldots, y_n$ separate into two classes. The first class includes all those values,—say $y_1, y_2, \ldots, y_r$,—for which it is possible to find a path that will convert y_1 into any other of the set; the second class includes the remaining roots which cannot be derived in this way from y_1. The r values $y_1, y_2, \ldots, y_r$ may be permuted by the description of a circuit; this cannot happen unless the circuit includes one or more branch-points. Suppose, to fix ideas, that the circuit includes one branch-point, b say The circuit may be contracted into a *loop*, that is, into a path from x_0 to a point b followed by a small circle round b and the same path as before reversed in direction (fig. 71); and the final value of y will be the same after as before the contraction. Now if y take near b one of a system of cyclic values $y_1, y_2, \ldots, y_r$ (§ 160), these values will permute cyclically. But if it take near b a value which is not one of a cyclic system, the final value is the same as the initial value. Thus in any case the symmetric combinations of $y_1, y_2, \ldots, y_r$ are unchanged and therefore they are one-valued functions of x which must be rational functions since the singular points are finite in number and non-essential. It follows that $y_1, y_2, \ldots, y_r$ satisfy an algebraic equation $F_1(x, y) = 0$; and this implies that F

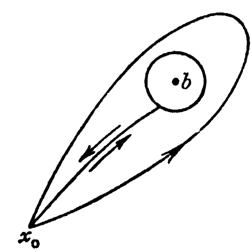


Fig. 71.

is divisible by F_1 contrary to the hypothesis that F is irreducible.

The equation $F(x, y) = 0$ *defines, then, a single function whose* n *values can be interchanged by choosing suitable circuits.*

Lastly any closed path whatever may be contracted into a

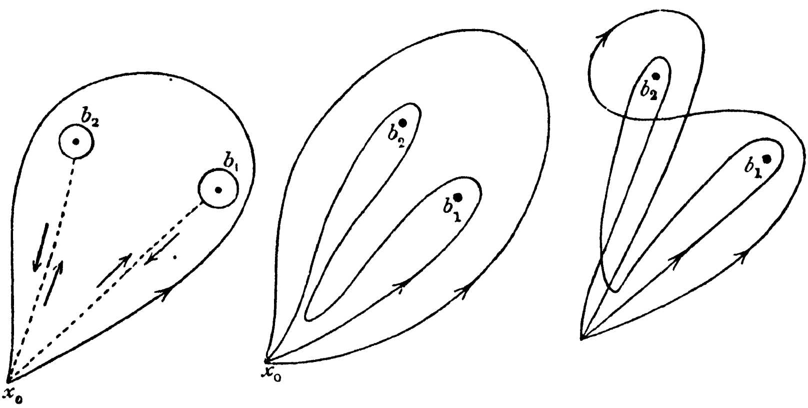


Fig. 72.

series of loops taken in a determinate order and described in a determinate sense; and the final value of y for that path may be inferred from the effect of the several loops taken in that order. Fig. 72 illustrates this deformation of closed paths; in each case the larger curve is deformed into the smaller (see also fig. 64).

162. Riemann Surface for an Algebraic Function. Without entering into details we shall indicate in a few words the nature of the Riemann surface that is used for the general algebraic function of x. We construct the surface on the x-plane. The surface contains n sheets because y is n-valued, and the n values of y are attached to the n points of the surface which lie vertically above the assigned point in the x-plane (supposed horizontal). These n sheets are connected by lines of passage which permit interchanges among the n values of y. If in the x-plane r values $y_1, y_2, \ldots, y_r$ permute cyclically when x describes a small circle (a), then on the surface the path cut out by a vertical cylinder standing on (a) is a spiral

which winds r times round (a) before returning to its initial place. As x passes r times from x_0 to x_0, the moving place (x, y) starts from (x_0, y_1) and passes successively to

$$(x_0, y_2), (x_0, y_3), \ldots, (x_0, y_r), (x_0, y_1).$$

The Riemann surface gives a clear idea of the extent of the circles of convergences of the series $P(x-x_0)$ at a non-singular point. These circles extend to the nearest singular places *in their respective sheets*; and here it is to be observed that a value $x=c$ which is an infinity or a branch-point of some of the n values of y is an ordinary point for other of these values, and that on the Riemann surface the places that correspond to these latter values offer no obstacles to the expansion of the corresponding circle of convergence.

CHAPTER XXII.

CAUCHY'S THEORY AND THE POTENTIAL.

163. Cauchy's Definition of a Monogenic Function. In this final chapter we shall point out some of the more salient features of Cauchy's definition of an analytic function, indicate the point of junction of the respective methods of Cauchy and Weierstrass, and then discuss a few simple cases in the theory of the potential for the plane.

Suppose that, following the plan of § 51, we assign in any way we please one real value to each point of a region Γ in the x-plane; in this way we construct for the region Γ a function of ξ, η. Let u, v be two functions so constructed; then $u + iv$, $= y$, is a function of x for the region Γ in the sense that when x is given, ξ and η and therefore u and v are determined. When u and v are left completely arbitrary, the combination of u, v into the single expression $u + iv$ offers no advantages; for, ultimately, u and v have to be considered separately as functions (in the sense of § 51) of the two independent variables ξ and η. Cauchy saw clearly the absolute necessity of sorting out from the total mass of functions of x those which may, in a useful sense, be regarded as functions not merely of ξ and η, but also of $\xi + i\eta$. He discarded such functions of x as $\xi - i\eta$, $\xi\eta$, and retained such functions as $(\xi + i\eta)^n$, $\sin(\xi + i\eta)$, etc. It is necessary to draw a line of division between the two classes; this can be done with the help of certain partial differential equations.

Take for $u+iv$ such expressions as $(\xi+i\eta)^n$, $\sin(\xi+i\eta)$, $e^{\xi+i\eta}$; then

$$\frac{\partial}{\partial\xi}(u+iv)=-i\frac{\partial}{\partial\eta}(u+iv) \quad \ldots\ldots\ldots\ldots\ldots(1);$$

and, on equating real and imaginary parts, (1) leads to

$$\frac{\partial u}{\partial\xi}=\frac{\partial v}{\partial\eta},\quad \frac{\partial u}{\partial\eta}=-\frac{\partial v}{\partial\xi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2).$$

On the other hand when $u+iv$ is equal to such an expression as $\xi-i\eta$,—one of the functions which Cauchy discarded,—these differential equations are not satisfied.

We have arrived at these equations by starting from definite expressions; we shall now show how they can be reached by more general reasoning.

Let u, v be defined as in the beginning of this article, and let

$$y=u+iv=fx,$$

where x is a selected point interior to Γ. When x increases by $\Delta x, =\Delta\xi+i\Delta\eta$, $x+\Delta x$ lying within Γ, let y increase by $\Delta y, =\Delta u+i\Delta v$. Now let Δx tend to zero; the question arises as to what conditions are necessary and sufficient in order that $\dfrac{\Delta y}{\Delta x}$ may tend to a *unique* finite limit when Δx tends to zero; when this limit exists at x it is denoted by $\dfrac{dy}{dx}$ or $f'x$ as in the case of the real variable, and called the derivate of x. Selecting an x in the finite part of the plane, $\sin x$ has a derivate $\cos x$ which does not depend on the way in which Δx tends to zero; on the other hand if we put $y=\xi-i\eta$ and allow $x+\Delta x$ to approach x along a ray through x with the chief amplitude α, we have for this mode of approach,

$$\lim_{\Delta x=0}\frac{\Delta y}{\Delta x}=\lim_{\Delta\xi=0,\,\Delta\eta=0}\frac{\Delta\xi-i\Delta\eta}{\Delta\xi+i\Delta\eta}=\frac{1-i\tan\alpha}{1+i\tan\alpha},$$

a quantity which varies with α, and therefore with the ray. Here there is a definite limit for the assigned mode of approach, but this limit depends on α and therefore there is not a unique limit for all modes of approach of Δx to zero. We shall find that the differential equations (2) are necessary conditions for the uniqueness of the limit at the selected point x.

I. *Necessary conditions for the uniqueness of the limit.* Evidently if there is to be a unique finite limit for $\frac{\Delta y}{\Delta x}$, this limit must be equal to the limit under the special circumstances,

(i) Δx constantly real and equal to $\Delta\xi$,

(ii) Δx constantly purely imaginary and equal to $i\Delta\eta$.

In case (i) we have

$$\frac{\Delta y}{\Delta x}=\frac{u(\xi+\Delta\xi,\eta)-u(\xi,\eta)}{\Delta\xi}+i\frac{v(\xi+\Delta\xi,\eta)-v(\xi,\eta)}{\Delta\xi};$$

it is therefore necessary that $\frac{dy}{dx}$ shall satisfy

$$\frac{dy}{dx}=\frac{\partial u}{\partial\xi}+i\frac{\partial v}{\partial\xi}\quad\text{.......................(3)},$$

while in the second case we have

$$\frac{\Delta y}{\Delta x}=\frac{u(\xi,\eta+\Delta\eta)-u(\xi,\eta)}{i\Delta\eta}+i\frac{v(\xi,\eta+\Delta\eta)-v(\xi,\eta)}{i\Delta\eta},$$

so that

$$\frac{dy}{dx}=-i\frac{\partial u}{\partial\eta}+\frac{\partial v}{\partial\eta}\quad\text{.......................(4)}.$$

Equating (3) and (4) we get the differential equations (2); these differential equations require, by implication, that u, v shall have partial derivates with respect to ξ and η at the selected point x.

Hitherto we have considered only two special modes of approach of Δx to zero. If there is to be a unique and finite limit for $\frac{\Delta y}{\Delta x}$ for *all* modes of approach, then we must have (replacing $x+\Delta x$ by x' for shortness),

$$\frac{dy}{dx}=\lim_{x'=x}\frac{fx'-fx}{x'-x}=f'x=\frac{\partial u}{\partial\xi}+i\frac{\partial v}{\partial\xi}\quad\text{...........(5)},$$

at the selected point x. It is only another way of stating (5) if we assert that there is a derivate at the point x provided a positive number δ can be found such that

$$\left|\frac{fx'-fx}{x'-x}-f'x\right|<\epsilon\quad\text{........................(6)},$$

for all values of x' that satisfy the inequality $|x'-x|<\delta$. The inequality (6) shows that

$$fx'-fx=(x'-x)\left[f'x+r(x'-x)\right]\quad\text{...........(7)},$$

where $r(x'-x)$ is a function of $x'-x$ such that $\lim\limits_{x'=x} r(x'-x)=0$.
Replacing $fx'-fx$ by $\Delta u+i\Delta v$, $\Delta\xi$ and $\Delta\eta$ by h and k, equating real and imaginary parts, and employing the equations (2), we see that

$$\left.\begin{aligned} \Delta u &= h\frac{\partial u}{\partial \xi}+k\frac{\partial u}{\partial \eta}+hr_1+kr_2, \\ \Delta v &= h\frac{\partial v}{\partial \xi}+k\frac{\partial v}{\partial \eta}+hr_3+kr_4, \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots(8),$$

where r_1, r_2, r_3, r_4 are four functions of h, k such that

$$\lim_{h=0,\,k=0} r_\lambda = 0, \qquad \lambda = 1, 2, 3, 4.$$

In order, then, that fx may have a derivate at a selected point x, it is necessary that

(1) *u, v shall satisfy the differential equations*

$$\frac{\partial u}{\partial \xi}=\frac{\partial v}{\partial \eta},\quad \frac{\partial u}{\partial \eta}=-\frac{\partial v}{\partial \xi};$$

(2) *u, v shall be such that Δu, Δv shall be expressible in the forms*

$$\Delta u=\frac{\partial u}{\partial \xi}\Delta\xi+\frac{\partial u}{\partial \eta}\Delta\eta+r_1\Delta\xi+r_2\Delta\eta,$$

$$\Delta v=\frac{\partial v}{\partial \xi}\Delta\xi+\frac{\partial v}{\partial \eta}\Delta\eta+r_3\Delta\xi+r_4\Delta\eta,$$

where $\lim r_\lambda=0$ ($\lambda=1, 2, 3, 4$) when $\Delta\xi$, $\Delta\eta$ tend independently to zero.

Before passing on to consider whether these two conditions are also sufficient, we wish to point out a consequence of the equation (5); namely that $\dfrac{fx'-fx}{x'-x}$ converges *uniformly* to its limit $f'x$ when x' tends to x (§ 55), whatever be the mode of approach of x' to x; in particular this is true when the approach is along rays through x.

II. *Sufficient conditions for the uniqueness of the limit of* $\dfrac{\Delta y}{\Delta x}$.

The two conditions that we have just shown to be necessary if there is to be a unique and finite limit, are also sufficient. For,

using the second of these conditions and putting $h = \rho\cos\alpha$, $k = \rho\sin\alpha$, we see that

$$\frac{\Delta y}{\Delta x} = \frac{\dfrac{\partial u}{\partial\xi}\cos\alpha + \dfrac{\partial u}{\partial\eta}\sin\alpha + i\left(\dfrac{\partial v}{\partial\xi}\cos\alpha + \dfrac{\partial v}{\partial\eta}\sin\alpha\right)}{\cos\alpha + i\sin\alpha} + \psi(\rho, \alpha),$$

where $\psi(\rho, \alpha)$ is a function of ρ and α such that $\lim_{\rho=0}\psi(\rho, \alpha) = 0$. The expression on the right-hand side tends to the limit

$$\frac{\left(\dfrac{\partial u}{\partial\xi} + i\dfrac{\partial v}{\partial\xi}\right)\cos\alpha + \left(\dfrac{\partial u}{\partial\eta} + i\dfrac{\partial v}{\partial\eta}\right)\sin\alpha}{\cos\alpha + i\sin\alpha}$$

or $\dfrac{\partial u}{\partial\xi} + i\dfrac{\partial v}{\partial\xi}$, since $\cos\alpha + i\sin\alpha$ is a factor of the numerator by reason of equation (1). Thus $\dfrac{dy}{dx}$ exists and is equal to $\dfrac{\partial u}{\partial\xi} + i\dfrac{\partial v}{\partial\xi}$.

Suppose now that it is given that u, v admit at x continuous first derivates with respect to ξ, η, then the second of the two conditions that we have been considering is necessarily satisfied and may be omitted. Hence *when the partial derivates of u, v of the first order are continuous functions of ξ, η at a selected point x, the differential equations* (2) *constitute the necessary and sufficient conditions for the existence of dy/dx at this point.* The continuity of fx is implied in the existence of $f'x$, as is seen by inspection of formula (7); the new conditions that have just been imposed make $f'x$ also continuous.

To indicate that the function $y, = fx$, has the property that $\dfrac{\Delta y}{\Delta x}$ tends, in general, to a unique finite limit Cauchy employed the term *monogenic*; in this way he excluded such functions as $\xi - i\eta$ as *non-monogenic.* Riemann dispensed with the adjective; in his terminology "a complex variable w is called a function of another complex variable z, when the former varies with the latter, in such a way that the value of the derivate dw/dz is independent of the value of the differential dz.*"

* Riemann, "Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse," *Ges. Werke*, p. 5.

164. Difficulties underlying Cauchy's definition. We shall now point out some of the difficulties that underlie Cauchy's definition of a monogenic function.

I. Is monogenic to mean monogenic at a point, or monogenic over a region? It is the usual, though not the invariable, custom to consider monogenic as an adjective applying to a region. Given that a function is monogenic over a region, is this to mean that the function has a finite derivate at all points of the region, or is allowance to be made for exceptional points, lines, regions? To take a simple example, is $1/x$ to be regarded as monogenic in a region which contains $x=0$? It is customary to make allowance for such points: e.g. $e^{1/x}$ may be treated as monogenic over the whole plane although it has no derivate at $x=0$.

To a considerable extent it is merely a matter of nomenclature whether we do or do not admit singular points into the regions over which functions are monogenic. But there are necessary restrictions to this practice. In § 94 we discussed an arithmetie expression which defines in different regions parts of distinct analytic functions: Weierstrass, in view of this fact, regarded such an expression not as defining a single monogenic function, but as defining parts of two distinct monogenic functions. In fact one of the lessons to be drawn from a study of analytic functions is the importance of keeping clearly distinct the idea of a function defined by properties and the idea of the dependence implied in an arithmetic expression; the two ideas are by no means necessarily coextensive. An expression may represent an analytic function only partially, e.g. $\sum_{n=0}^{\infty} x^n$; or again it may represent different analytic functions completely in different regions; or it may represent certain analytic functions partially in certain regions and other analytic functions completely in other regions. We shall not stop to justify these statements, though it is not difficult to do so with the aid of § 94. The point which we desire to emphasize is that Cauchy's definition implies in various ways a considerable preliminary grasp of the logical possibilities attached to the study of singular points.

II. The inequality (6) shows that $\dfrac{fx'-fx}{x'-x}$ is to converge *uniformly* to its limit (§ 55). Hence if x' tends to x along rays through x, it is necessary for the existence of $f'x$ that the convergence for these modes of approach shall be *uniform*. We shall show below that it is possible to construct a function fx which shall have a unique limit for $\dfrac{\Delta y}{\Delta x}$ for all modes of approach of x' to x along rays and yet not have a derivate $f'x$. From what has been said the convergence of $\dfrac{fx'-fx}{x'-x}$ must, in this case, be non-uniform for the system of rays. Here then we have a difficulty connected not with the singular but with the ordinary points of a function: the recognition, namely, that the convergence to the limit must be uniform. Riemann's definition of a function (§ 163) must not, therefore, be so interpreted as to restrict the paths of approach of Δx to zero to straight lines.

Stolz* illustrates the case of non-uniform convergence to the limit $f'x$ along rays through o, by the function fx which is equal to $2x\dfrac{\xi\eta^2}{\xi^2+\eta^4}$ for values of x different from o, and is equal to o for $x=0$. If we put $\xi=\rho\cos a$, $\eta=\rho\sin a$, where a is the chief amplitude of a ray through o, we have

$$\lim_{\rho=0}\frac{fx-f0}{x}=0,$$

so that for each ray the limit of $\dfrac{\Delta y}{\Delta x}$ exists and is equal to o; but the convergence to this limit o is not uniform. To convince ourselves of this fact let us put $\epsilon=$ a proper fraction, and take the system of rays from $a=0$ to $a=\pi/2$. Along each ray we shall suppose that a stroke of length δ' is measured off, where δ' is the greatest positive number such that we have, for all points x situated on the ray at a distance from o less than δ',

$$\left|\frac{fx-f0}{x}\right|,\quad =\frac{2\xi\eta^2}{\xi^2+\eta^4},\quad <\epsilon.$$

As a approaches nearer and nearer to $\pi/2$, δ' gets smaller and smaller; for the stroke cannot extend as far as the parabola $\eta^2=\xi$, because on this parabola the expression on the left-hand side of the inequality has the value 1; hence the lower limit of δ' is o, and there is no single value δ of the kind considered in connexion with the inequality (6) which will serve for all rays of the system. In other words $\dfrac{fx-f0}{x}$ does not converge uniformly to o,

* Stolz, *Grundzüge der Differential- und Integralrechnung*, vol. ii. p. 80.

o is not the limit (in the sense of equation (5)) of this ratio, and the function fx is not a monogenic function of x.

III. There are difficulties which relate to the postulation of continuity for the derivates of fx. Cauchy's theory of functions has in view the same functions as those considered by Weierstrass, namely, functions fx which are analytic about an ordinary point x. This being so, the derivates $f'x, f''x, \ldots$ must all exist and be continuous at x; but it is evidently undesirable to postulate explicitly the existence and continuity of $f''x$, $f'''x, \ldots$, if these facts are consequences of the existence and continuity of fx and $f'x$. Among Cauchy's many important contributions to mathematical knowledge, a high rank will always be assigned to his proof that a theory of functions can be constructed on the basis of monogeneity coupled with the continuity of $f'x$ (Cauchy himself included the continuity of $f'x$ in the definition of a monogenic function). To appreciate properly the remarkable nature of this discovery of Cauchy's it is necessary to bear in mind that when the variable is real a function (in the sense of § 48) may be continuous and yet not have a derivate, so that the existence of $f''x$, $f'''x, \ldots$ is by no means an evident consequence of the existence and continuity of fx and $f'x$.

Granting that at an ordinary point a monogenic function must have a continuous derivate if it is to be admitted into a theory of functions of a complex variable, the question at once arises:—does there exist a class of monogenic functions fx for which $f'x$ is discontinuous? To present the matter in a somewhat different way:—what is the irreducible minimum of conditions to be imposed upon fx, if we desire to have fx, in general, analytic about x? It may very possibly be true that in order that fx may be analytic about a point x it is necessary and sufficient that fx shall be one-valued, finite, and admit a finite derivate for all points of a neighbourhood of x*.

165. Extended form of Taylor's Theorem. Cauchy proved his theorem on integration (§ 118) for every function

* Pringsheim seems to assert this in *Math. Ann.* vol. xliv. p. 80 (1893), but see a foot-note in his paper on Cauchy's theorem, *Sitzber. d. k. bay. Ak. d. Wiss.* (1895), vol. xxv.

fx which is one-valued and admits a one-valued continuous derivate over a closed region Γ. The same consequences follow as before as regards residues, etc.; the only difference is that fx is defined in a new way. But that fx is analytic about each point of Γ now requires proof, whereas, before, this property was postulated for fx from the start. The process by which Cauchy established the theorem that *when fx is one-valued and admits a one-valued and continuous derivate at each point of a closed region (c, R), it can be expanded as a power series $P(x-c)$ whose radius of convergence is not less than R*, is precisely that used in § 122, the ring being supposed reduced to (c, R) by the vanishing of R'. The coefficients in the expansion

$$fx = a_0 + a_1(x-c) + a_2(x-c)^2 + \ldots + a_n(x-c)^n + \ldots,$$

are the definite integrals $\dfrac{1}{2\pi i}\displaystyle\int_C \frac{fx\,dx}{(x-c)^{n+1}}$, so that

$$fc = \frac{1}{2\pi i}\int_C \frac{fx\,dx}{x-c}, \quad f^{(n)}c = \frac{1}{2\pi i}\int_C \frac{fx\,dx}{(x-c)^{n+1}}.$$

It is evident from what has been said that the theorem which we have called Cauchy's theorem (§ 118) occupies a central position in the theory of analytic functions as developed by means of integration. We shall give at the end of this chapter a second proof of this theorem which involves the transformation of a double into a simple integral by means of Green's theorem. This proof was discovered by Cauchy and was used later by Riemann.

By proving that a function fx which is one-valued, continuous, and admits a continuous derivate over a neighbourhood of x, is analytic about x we have arrived at the point of junction of the theories of Cauchy and Weierstrass. The analytic function is always monogenic; but to be able to say conversely that Cauchy's monogenic function is always analytic the idea 'monogenic' must be made more precise in that part which Cauchy left vague. One analytic expression must be allowed in some cases to represent more than one function completely or partially. This is more an addition to than a modification of Cauchy's idea.

Cauchy's theorem was contained in germ in a memoir on definite integrals (1814); explicitly in a supplement to this memoir (1825). Gauss enunciated the theorem in 1812 in a private letter to Bessel. Green's theorem is contained in his *Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*, 1828; that proof of Cauchy's theorem (§ 170) which depends on Green's theorem was published by Cauchy in 1846. It is possible, but not probable, that Cauchy was acquainted with Green's memoir, for it was not till 1850 that Green's memoir became readily accessible to mathematicians. Riemann's proof, which applies also to many-valued functions, was published in his inaugural dissertation (1851), *Ges. Werke.* p. 12; it agrees essentially with Cauchy's.

166. The Potential. When $u+iv$ is an analytic function of x, say

$$u+iv=f(\xi+i\eta),$$

then

$$\frac{\partial^2 u}{\partial \xi^2}+i\frac{\partial^2 v}{\partial \xi^2}=f''(\xi+i\eta),$$

and

$$\frac{\partial^2 u}{\partial \eta^2}+i\frac{\partial^2 v}{\partial \eta^2}=-f''(\xi+i\eta),$$

the second derivate f'' being definite at any point about which f is analytic.

Hence adding we have

$$\frac{\partial^2 u}{\partial \xi^2}+\frac{\partial^2 u}{\partial \eta^2}=0, \quad \frac{\partial^2 v}{\partial \xi^2}+\frac{\partial^2 v}{\partial \eta^2}=0.$$

Thus u and v satisfy the same differential equation, which is Laplace's equation for two dimensions.

Thus our theory applies to an important physical concept, namely the *potential in a plane.* By this term we mean any function of ξ, η which can be used as the u or v of an analytic function $u+iv$ of $\xi+i\eta$. It partakes of the properties of the associated analytic function; its singular points are to be sought among the singular points of the analytic function, and it may be many-valued. We shall confine our attention however to a region of the plane, and within that region we shall suppose the potential to be one-valued and continuous; then at all points of the region it will have continuous derivates of all orders. Historically the developments of the potential, as a measurable physical concept (like a distance or an angle) and as a function

of the position in the plane (that is, ultimately, as a number), are closely connected.

An introductory discussion of the potential in a plane will now be given; for its appropriate application and extension in the various departments of physics reference must of course be made to the physical treatises.

When we write down an analytic function of $\xi+i\eta$ or x, and separate the real part and the imaginary part so as to express the function in the form $u+iv$, where u and v are real, we obtain a pair of potentials u and v. A convenient way to think of such a function as u or v is to conceive an ordinate to the x-plane erected at the point (ξ, η) or x, the length of the ordinate representing the value of the function u or v. A characteristic property of the surface generated by the extremity of the ordinate is that the projection on the x-plane of the indicatrix at a point about which $u+iv$ is analytic, is a rectangular hyperbola. For we have, by Taylor's theorem,

$$u'-u=\frac{\partial u}{\partial \xi}(\xi'-\xi)+\frac{\partial u}{\partial \eta}(\eta'-\eta)+\frac{\partial^2 u}{\partial \xi^2}(\xi'-\xi)^2/2!$$
$$+\frac{\partial^2 u}{\partial \xi \partial \eta}(\xi'-\xi)(\eta'-\eta)+\frac{\partial^2 u}{\partial \eta^2}(\eta'-\eta)^2/2!+\ldots;$$

the projection of the indicatrix is obtained by keeping $u'-u$ constant and omitting powers above the second; and the condition that the cylinder so obtained shall have as cross-section a rectangular hyperbola is $\frac{\partial^2 u}{\partial \xi^2}+\frac{\partial^2 u}{\partial \eta^2}=0$.

Potentials occur in pairs, u and v, where $u+iv$ is an analytic function of $\xi+i\eta$. The two are often said to be *conjugate.* It is to be observed that if v is the conjugate of u, then not u but $-u$ is the conjugate of v, so that the choice of the adjective conjugate is not entirely happy. When the one potential is given the other is determined, save as to a constant, by the equations

$$\frac{\partial u}{\partial \xi}=\frac{\partial v}{\partial \eta}, \quad \frac{\partial u}{\partial \eta}=-\frac{\partial v}{\partial \xi};$$

but it is usually more important to know the form of the analytic function than to know u or v separately.

Some simple potentials are:

(1) The distance from a point to a fixed line; for evidently ξ and η are themselves potentials and so is $\alpha\xi+\beta\eta+\gamma$ where α, β, γ are constant. It must be noticed that because Laplace's equation is linear, the sum of any given number of potentials is itself a potential.

The representative surface is here any plane.

(2) The amplitude of x. For $\log x = \operatorname{Log}\rho + i\theta$, whence $\operatorname{Log}\rho$ and θ are potentials.

Thus the angle made by the line from any fixed point to x with any selected zero of direction is a potential. The representative surface is a helicoid.

The potential includes then two of the fundamental quantities of geometry, distance of point and line, and angle. But it does not include the distance ρ of two points, nor any power of this distance; ρ^n is not a potential, but $\operatorname{Log}\rho$ is; whereas in three dimensions $1/\rho$ is a potential. Hence the present potential is often called *logarithmic.*

(3) Let $\theta_1, \theta_2, \ldots, \theta_n$ be the amplitudes of strokes from fixed points to x. Then

$$\alpha_1\theta_1+\alpha_2\theta_2+\ldots+\alpha_n\theta_n$$

is a potential.

167. The Equipotential Problem. The problem of potentials is: given a system of real values continuous, in general, along a circuit, to find a potential u which shall be one-valued and continuous within the region bounded by this circuit and shall take the assigned values on the contour. A case of special simplicity and importance is when the boundary is made up of equipotential lines, that is to say, lines along each of which u is a constant. We may call this the equipotential problem; and we shall assume here that the solution is unique. The uniqueness of the solution is strongly suggested by physical considerations; and is readily proved by means of Green's theorem (§ 169).

But even when the boundary consists of equipotentials the problem in its generality is quite beyond our scope.

The inverse process, however, of determining the region when we start with a known analytic function and therefore with a known potential, is merely a matter of mapping.

For example, as a case of (3) of § 166, let the region be the upper half of the x-plane, and let θ_1, θ_2 be the amplitudes of strokes from fixed points of the real axis, then $\theta_1 - \theta_2$ is a potential; it takes along the real axis the values $0, \pi, 0$. Hence, conversely, if the real axis is divided into three parts which are kept at potentials $0, \pi, 0$, the potential at any point of the upper half-plane is $\theta_1 - \theta_2$. The equipotential lines are

$$\theta_1 - \theta_2 = \text{constant},$$

and are therefore arcs of circles.

The conjugate potential is $\operatorname{Log} \rho_1 - \operatorname{Log} \rho_2$; it is constant along the circles $\rho_1/\rho_2 = \text{constant}$. That these circles cut the arcs at right angles appeared in § 21; but it is an illustration of the general property of isogonality, by which the orthogonal straight lines $u = \text{constant}$, $v = \text{constant}$, in the $(u + iv)$-plane map into orthogonal lines, provided we have at their intersection

$$u + iv - u_0 - iv_0 = P_1(x - x_0).$$

If we have $P_n(x - x_0)$ instead of $P_1(x - x_0)$, then (§ 108) the angle in the $(u + iv)$-plane is $n \times$ (the angle in the x-plane).

When u is the potential considered, the lines $v = \text{constant}$ in the x-plane are the lines of flow (lines of force in Electrostatics); conversely, when v is the potential considered, the lines of flow become equipotentials, and vice versa.

The present example solves the equipotential problem for a crescent, or region bounded by two circular arcs when these arcs are kept at given constant potentials. For, measuring the amplitudes θ_1, θ_2 from the intersections of the arcs, the equations of the arcs are

$$\theta_1 - \theta_2 = \gamma_1,$$

$$\theta_1 - \theta_2 = \gamma_2;$$

and the function $\alpha(\theta_1 - \theta_2) + \beta$ is a potential which takes

constant values u_1, u_2 along these arcs. The constants α and β are found at once when these constant values are given.

Thus we have for any coaxial arcs for which

$$\theta_1 - \theta_2 = \gamma,\ \gamma_1,\ \gamma_2,$$
$$u = \alpha\gamma + \beta,$$
$$u_1 = \alpha\gamma_1 + \beta,$$
$$u_2 = \alpha\gamma_2 + \beta,$$

and

$$\begin{vmatrix} u & u_1 & u_2 \\ \gamma & \gamma_1 & \gamma_2 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

The potential $\alpha_0 + \Sigma\alpha_m\theta_m$ applies similarly to the case when a straight line, say the real axis, is divided into $n+1$ intervals, each kept at constant potential. Let the first interval be that for which all the θ's are zero, the second that for which $\theta_1 = \pi$ and the remaining θ's are zero, and so on; also let $v_0, v_1, v_2, \ldots, v_n$ be the $n+1$ values of the potential v along the 1st, 2nd, 3rd, ..., $(n+1)$th intervals. For simplicity let the straight line be the real axis and let the region bounded by this straight line be the upper half of the x-plane. We shall have

$$\alpha_0 = v_0,$$
$$\alpha_0 + \pi\alpha_1 = v_1,$$
$$\alpha_0 + \pi(\alpha_1 + \alpha_2) = v_2,$$

and so on, $n+1$ equations to determine the constants.

Let us suppose (fig. 73) that the potentials $v_0, v_1, v_2, \ldots, v_n$ are in decreasing order; it follows that $\alpha_1, \alpha_2, \ldots, \alpha_n$ are negative. The figure formed in the plane of $u + iv$ must be noticed. It is of course formed by a series of lines $v = \text{constant}$; but of these lines how much is to be taken?

The analytic function $u + iv$ is here

$$y = i\alpha_0 + \beta_0 + \Sigma\alpha_m \log(x - \xi_m),$$

where ξ_m is a dividing point of the straight line, which is taken as the real axis in the x-plane. And the conjugate potential to v is

$$u = \beta_0 + \Sigma\alpha_m \operatorname{Log} \rho_m,$$

where $\rho_m = |\xi - \xi_m|$.

Hence since α_m is negative, as ξ describes the real axis from right to left u passes from $-\infty$ to $+\infty$ at ξ_1, from $+\infty$ at ξ_1 to $+\infty$ at ξ_2, and so on; finally passing back to $-\infty$ only when ξ has passed ξ_n. In fact u attains a minimum when

$$\Sigma\alpha_m/(\xi - \xi_m) = 0,$$

an equation with $n - 1$ roots separating the points ξ_m. Thus the map in the y-plane is as shown in fig. 73 (which is drawn for the case $n = 4$), and contains a system of straight lines which are parallel to the real axis and which extend to ∞ on the right, but not to $-\infty$ on the left except in the case of the first and last lines. Each of the lines intermediate between the first and last is traversed twice, and the terminal points of these lines can be found from the roots of the above equation. The straight lines p_0p_1, p_1p_2, p_2p_3, p_3p_4, are the maps of the small semi-circles (assumed extremely small); the dotted line on the left is the map of the dotted semi-circle (assumed extremely large) in the x-plane by which the x-circuit is completed

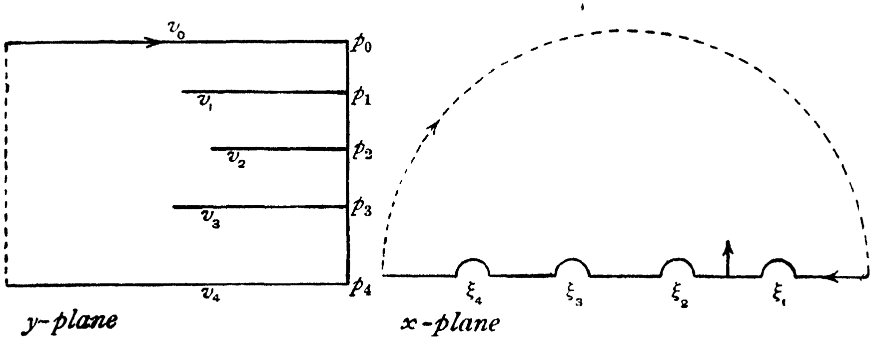


Fig. 73.

Ex. Consider the case when the given potentials $v_0, v_1, v_2, \ldots, v_n$ are not in order of magnitude.

We can solve in the same way the equipotential problem for the case of a circular boundary divided into $n + 1$ intervals, or we can infer this from the straight line by the principle of inversion. For if

$$y = u + iv = fx.$$

and we write $x = (ax' + b)/(cx' + d)$, then x and x' describe inverse curves X and X'; and if the function fx has been

determined which maps the parallel half-lines $u = \text{constant}$ on X, then the function $f\{(ax' + b)/(cx' + d)\}$ will map the same half-lines on X'.

With regard to equipotential lines in general it must be noticed that at a point of the x-plane at which v is infinite u may have any value, so that all the lines $u = \text{constant}$ will lead to the infinity.

168. Schwarz's and Christoffel's mapping of a straight Line on a Polygon. Suppose that we have a rectilinear polygon in the x-plane with exterior angles $\alpha_1\pi, \alpha_2\pi, \ldots, \alpha_n\pi$, measured in the negative sense so that $\Sigma\alpha_m = -2$. The polygon is supposed convex so that $0 > \alpha_m > -1$.

If with these conditions we write

$$z = \exp y,$$

then

$$z = \exp(\beta_0 + i\alpha_0)\,\Pi\,(x - \xi_m)^{\alpha_m},$$

x and y being connected by the equation (1) of the last article.

The parallel half-lines in the y-plane become rays in the z-plane from ∞ towards the origin, but only the first and last reach the origin. Moreover since for large values of x, $z = P_2(1/x)$

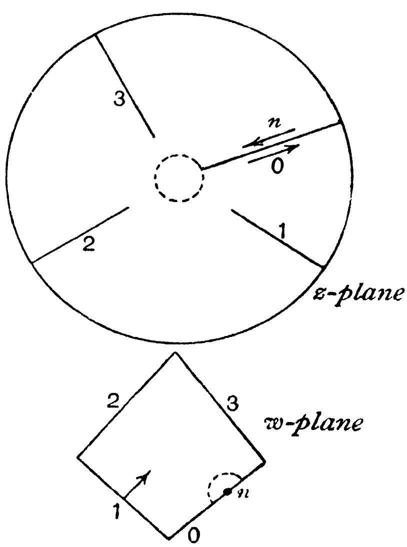


Fig. 74.

since $\Sigma\alpha_m = -2$, the semi-circle of fig. 73 becomes a complete circle; that is, the first and last lines of the z-plane coincide.

Now let $w = \int z\,dx$; then as x moves along its real axis from $+\infty$ to 0, we have $dw = z dx$, and

$$\begin{aligned}\text{am}\, dw &= \text{am}\, z + \text{am}\, dx \\ &= \text{am}\, z + \pi.\end{aligned}$$

Hence the straight lines in the z-plane of fig. 74 (drawn for $n=4$, to fix ideas) remain straight lines in the w-plane; and since, near ξ_m,

$$w = \int (x-\xi_m)^{\alpha_m} P_0(x-\xi_m)\,dx = (x-\xi_m)^{1+\alpha_m} Q_0(x-\xi_m),$$

there is no abrupt change in w when x describes the small semi-circle round ξ_m; in the limit the map in the w-plane is a polygon; and since near $x = \infty$

$$w = \int P_2(1/x)\,dx = P_1(1/x),$$

the polygon is closed.

Thus, omitting the constant $\exp(\alpha_0 + i\beta_0)$ which merely displaces the polygon, we have

$$w = \int \Pi\,(x-\xi_m)^{\alpha_m}\,dx,$$

as the equation which maps the ξ-axis on the polygon. By the property of isogonality, when x turns to the right into the upper half-plane, as indicated by the arrow in the right-hand part of fig. 73, w turns also to the right into the interior of the polygon, as indicated by the arrow in the w-polygon of fig. 74. Hence the upper half of the x-plane maps into the interior of the w-polygon.

Ex. Let $w = \int dx/\sqrt{1-x^2}$ and let x describe the real axis. Draw the path of w.

169. Green's Theorem for Two Dimensions. Let u and v be two potentials; let C be a circuit within which there are no singular points of the analytic function $u+iv$. Then, within C, u, v and their derivates of the first and second orders are continuous. The simplest supposition with respect to C is that it is an oval (for example an ellipse) which is met by a straight line parallel to the axis of ξ or η in not more than two real points.

Let us assume, for simplicity, that C has this form. We propose to transform

$$\iint_\Gamma \left(\frac{\partial u}{\partial \xi}\frac{\partial v}{\partial \xi} + \frac{\partial u}{\partial \eta}\frac{\partial v}{\partial \eta}\right) d\xi\, d\eta,$$

taken over the region Γ within C, into a simple integral taken over C itself in the positive direction (§ 108).

Since $$\frac{\partial u}{\partial \xi}\frac{\partial v}{\partial \xi} = \frac{\partial}{\partial \xi}\left(u\frac{\partial v}{\partial \xi}\right) - u\frac{\partial^2 v}{\partial \xi^2},$$
it follows that

$$\iint_\Gamma \frac{\partial u}{\partial \xi}\frac{\partial v}{\partial \xi}\, d\xi\, d\eta = \int_C u\frac{\partial v}{\partial \xi}\, d\eta - \iint_\Gamma u\frac{\partial^2 v}{\partial \xi^2}\, d\xi\, d\eta \text{ ...... (1).}$$

For $\iint \frac{\partial}{\partial \xi}\left(u\frac{\partial v}{\partial \xi}\right) d\xi d\eta$ over the shaded strip of thickness k is equal to $$k\int_{\xi_0}^{\xi_1} \frac{\partial}{\partial \xi}\left(u\frac{\partial v}{\partial \xi}\right) d\xi,$$

and therefore to $k\left[u\frac{\partial v}{\partial \xi}\right]$ at p, $-k\left[u\frac{\partial v}{\partial \xi}\right]$ at q. Remarking that pp', $q'q$ are positive arcs, we see that $k = d\eta$ at p and $-d\eta$ at q and hence the part due to the strip is the sum of elements $u\frac{\partial v}{\partial \xi}d\eta$ at p and q, and the part due to all the strips, that is to Γ, is $\int_C u\frac{\partial v}{\partial \xi}\, d\eta$, which is what we wished to prove.

Applying similar reasoning to $\iint \frac{\partial u}{\partial \eta}\frac{\partial v}{\partial \eta}\, d\xi\, d\eta$ we get

$$\iint_\Gamma \frac{\partial u}{\partial \eta}\frac{\partial v}{\partial \eta}\, d\xi\, d\eta = \int_C v\frac{\partial u}{\partial \eta}\, d\xi - \iint_\Gamma v\frac{\partial^2 u}{\partial \eta^2}\, d\xi\, d\eta \text{ ....(2).}$$

By adding (1) and (2),

$$\iint_\Gamma \left(\frac{\partial u}{\partial \xi}\frac{\partial v}{\partial \xi} + \frac{\partial u}{\partial \eta}\frac{\partial v}{\partial \eta}\right) d\xi\, d\eta = \int_C u\left(\frac{\partial v}{\partial \xi}\, d\eta - \frac{\partial v}{\partial \eta}\, d\xi\right)$$
$$- \iint_\Gamma u\nabla^2 v\, d\xi\, d\eta \text{......(3),}$$

where ∇^2 denotes $\frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2}$.

This formula (3) is one of the forms of Green's theorem; another form can be found by interchanging u, v in (3) and equating the expressions on the right-hand sides of (3) and of the new formula.

Let s be the length of the arc measured along C in the positive sense from a point of reference to the variable point p, and at p let a normal be drawn inwards of length n. The coordinates x, y of the extremity of this normal can be expressed (at least theoretically) in terms of s, n. Then

$$\frac{\partial \xi}{\partial n} = \cos \psi, \quad \frac{\partial \eta}{\partial n} = \sin \psi, \quad \frac{\partial \xi}{\partial s} = \sin \psi, \quad \frac{\partial \eta}{\partial s} = \cos \psi,$$

where ψ is the angle made by the tangent at p with the positive direction of the axis of ξ. Hence

$$\frac{\partial \xi}{\partial s} = \frac{\partial \eta}{\partial n}, \quad \frac{\partial \eta}{\partial s} = -\frac{\partial \xi}{\partial n},$$

and (3) becomes

$$\iint_\Gamma \left(\frac{\partial u}{\partial \xi}\frac{\partial v}{\partial \xi} + \frac{\partial u}{\partial \eta}\frac{\partial v}{\partial \eta}\right) d\xi\, d\eta = -\int_C u \frac{\partial v}{\partial n} ds$$

$$- \iint_\Gamma u\nabla^2 v\, d\xi\, d\eta \;\ldots\ldots(4).$$

Interchanging u and v and subtracting the result from (4) we have the standard form for Green's theorem :—

$$\int_C \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n}\right) ds + \iint_\Gamma (u\nabla^2 v - v\nabla^2 u)\, d\xi\, d\eta = 0 \ldots\ldots(5).$$

170. Cauchy's Theorem. A similar method of transformation of a double into a single integral when applied to

$$-\iint_\Gamma \left(\frac{\partial u}{\partial \eta} + \frac{\partial v}{\partial \xi}\right) d\xi\, d\eta, \quad \iint_\Gamma \left(\frac{\partial u}{\partial \xi} - \frac{\partial v}{\partial \eta}\right) d\xi\, d\eta,$$

converts these into

$$\int_C (u d\xi - v d\eta), \quad \int_C (v d\xi + u d\eta),$$

and hence

$$\int_C (u + iv)\, dx = \int_C (u d\xi - v d\eta) + i \int_C (v d\xi + u d\eta)$$

$$= -\iint \left(\frac{\partial u}{\partial \eta} + \frac{\partial v}{\partial \xi}\right) d\xi\, d\eta + \iint \left(\frac{\partial u}{\partial \xi} - \frac{\partial v}{\partial \eta}\right) d\xi\, d\eta.$$

If the potentials u, v are conjugate, the double integrals vanish by reason of the equations satisfied by the conjugate potentials u, v; and hence Cauchy's theorem is proved for the circuit C.

Cauchy's integral $\frac{1}{2\pi i}\int fx\, dx/(x - c)$, see § 120, can also be

connected with the theory of the potential. The formula (5) was proved without making any use of the hypothesis that u and v are conjugate potentials; the theorems (1) to (5) are true if we merely postulate that u, v are to be one-valued and continuous within C together with their derivates of the first and second orders. Hence it is not necessary, even when u, v are solutions of Laplace's equation, that they should be conjugate. We may, in particular, write $u = 1$ in (5) and deduce the equation

$$\int_C \frac{\partial v}{\partial n} ds = -\iint_\Gamma \nabla^2 v \, d\xi \, d\eta = 0 \quad \ldots\ldots\ldots\ldots\ldots(6).$$

The restriction on C in equations (1) to (6) can be removed; moreover the boundary of Γ can, if we choose, be supposed to consist of several circuits instead of one; the simple integral is then taken positively with respect to Γ over all these circuits. As a special application take a region Γ bounded by a circuit C and a small circle (c, ϵ) about a point c within C; and let $u = \operatorname{Log} \rho$ where ρ is the distance of c from a variable point on C. Then $\nabla^2 u = 0$, $\nabla^2 v = 0$, and (5) becomes

$$\int_C \left(\operatorname{Log} \rho \frac{\partial v}{\partial n} - v \frac{\partial \operatorname{Log} \rho}{\partial n}\right) ds$$
$$= \int_{(c,\,\epsilon)} \left(\operatorname{Log} \rho \frac{\partial v}{\partial n} - v \frac{\partial \operatorname{Log} \rho}{\partial n}\right) ds \ \ldots(7).$$

But $\dfrac{\partial \operatorname{Log} \rho}{\partial n}, = \dfrac{1}{\rho}\dfrac{\partial \rho}{\partial n}$, has the value $-\dfrac{1}{\epsilon}$ along (c, ϵ); hence

$$\int_C \left(\operatorname{Log} \rho \frac{\partial v}{\partial n} - v \frac{\partial \operatorname{Log} \rho}{\partial n}\right) ds = \operatorname{Log} \epsilon \int_{(c,\,\epsilon)} \frac{\partial v}{\partial n} ds + \frac{1}{\epsilon} \int_{(c,\,\epsilon)} v ds$$
$$= \frac{1}{\epsilon} \int_{(c,\,\epsilon)} v ds, \text{ by (6);}$$

(c, ϵ) being now described positively with respect to c.

As the expression on the left does not depend on the value of ϵ,—for neither the integrand nor C depends on ϵ,—the integral on the right must be independent of ϵ; hence to evaluate it we make ϵ tend to zero and observe that $\dfrac{1}{\epsilon}\displaystyle\int_{(c,\,\epsilon)} v ds$ can be made to differ by as little as we please from $\dfrac{1}{\epsilon} v(z) \displaystyle\int_{(c,\,\epsilon)} ds$,

that is from $2\pi v(c)$. As the integral is known to be independent of ϵ, $2\pi v(c)$ must be its exact value. Thus

$$v(c) = \frac{1}{2\pi}\int_C \left\{\operatorname{Log} \rho \frac{\partial v}{\partial n} - v \frac{\partial \operatorname{Log} \rho}{\partial n}\right\} ds \quad \text{.........(8).}$$

The formula (8) is a particular case of the theorem

$$fc = \frac{1}{2\pi i}\int_C \frac{fx\, dx}{x - c}.$$

To prove this it is necessary to resolve both sides into the real and imaginary parts and then equate the imaginary parts; for the details of the verification we must refer the reader to Picard's *Traité d'Analyse*, vol. ii. pp. 109, 110. Even without this verification the dependence of fc,—the value of fx at an interior point,—on the values of the same function along the rim C suggests strongly a connexion between this formula of Cauchy's and the solution of the physical problem of the determination of the state of a body at an interior point when the boundary conditions are given.

LIST OF BOOKS.

General Treatises: Among elementary books dealing generally with the Theory of Functions may be mentioned BURKHARDT, Einführung in die Theorie der analytischen Functionen, and THOMÆ, Theorie der analytischen Functionen.

Students who wish for a fuller treatment should read FORSYTH's admirably written book, Theory of Functions of a complex variable. Copious references are there given to the original sources. Other general accounts are contained in our Treatise on the Theory of Functions, in PICARD's Traité d'Analyse, vols. i. and ii., and in JORDAN's Cours d'Analyse.

The following short list of books dealing with special departments of the subject may also be useful.

The notions of number, limit, etc.: CHRYSTAL, Algebra; STOLZ, Vorlesungen über allgemeine Arithmetik; TANNERY, Théorie des Fonctions d'une Variable réelle.

Trigonometry: HOBSON, Plane Trigonometry.

Elliptic Functions; APPELL and LACOUR, Principes de la Théorie des Fonctions elliptiques; TANNERY and MOLK, Éléments de la Théorie des Fonctions elliptiques; ENNEPER-MÜLLER, Elliptische Functionen.

Algebraic Functions and Abelian Integrals: APPELL and GOURSAT, Théorie des Fonctions algébriques et de leurs Intégrales; BAKER, On Abel's Theorem and the allied Theory; BRILL and NÖTHER, Die Entwickelung der Theorie der algebraischen Functionen, in the Jahresbericht der Deutschen Mathematiker-Vereinigung, 1894; KLEIN, On Riemann's Theory of algebraic Functions and their Integrals (translated by Miss Hardcastle); NEUMANN, Vorlesungen über Riemann's Theorie der Abel'schen Integralen; STAHL, Theorie der Abel'schen Functionen.

INDEX.

The numbers refer to the pages.

Printed in Austria.